



TESIS - TE142599

**PREDIKSI TEMUAN DENGAN METODE  
TRANSFORMASI MASALAH:  
PADA KASUS PEMERIKSAAN LAPORAN KEUANGAN WILAYAH  
PROVINSI KALIMANTAN TENGAH**

ALLANTUTRA GUSLAWA  
NRP 07111650067003

DOSEN PEMBIMBING  
Dr. Ir. Endroyono, DEA  
Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA  
KONSENTRASI PENGELOLA TIK PEMERINTAHAN (PETIK)  
DEPARTEMEN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI ELEKTRO  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2018





TESIS - TE142599

**PREDIKSI TEMUAN DENGAN METODE  
TRANSFORMASI MASALAH:  
PADA KASUS PEMERIKSAAN LAPORAN KEUANGAN WILAYAH  
PROVINSI KALIMANTAN TENGAH**

ALLANTUTRA GUSLAWA  
NRP 07111650067003

DOSEN PEMBIMBING  
Dr. Ir. Endroyono, DEA  
Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA  
KONSENTRASI PENGELOLA TIK PEMERINTAHAN (PETIK)  
DEPARTEMEN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI ELEKTRO  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2018



## LEMBAR PENGESAHAN


Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Teknik (M.T.)  
di  
Institut Teknologi Sepuluh Nopember

oleh:


Allantutra Guslaw  
NRP. 07111650067003

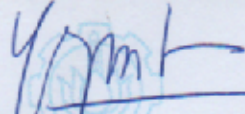
Tanggal Ujian : 06 Juli 2018  
Periode Wisuda : September 2018

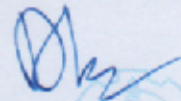
Disetujui oleh:

  
1. Dr. Ir. Endroyono, DEA  
NIP. 196504041991021001

(Pembimbing I)

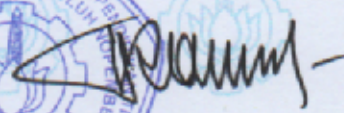
  
2. Dr. Supeno Mardi Susiki Nugroho, S.T., M.T. (Pembimbing II)  
NIP. 197003131995121001

  
3. Prof. Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc. (Penguji)  
NIP. 195409251978031001

  
4. Dr. Adhi Dharma Wibawa, S.T., M.T. (Penguji)  
NIP. 197605052008121003

(Penguji)

Dekan Fakultas Teknologi Elektro

  
Dr. Tri Arief Sardjono, S.T., M.T.  
NIP. 197002121995121001

*Halaman ini sengaja dikosongkan*



## **PERNYATAAN KEASLIAN TESIS**

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul “**PREDIKSI TEMUAN DENGAN METODE TRANSFORMASI MASALAH: PADA KASUS PEMERIKSAAN LAPORAN KEUANGAN WILAYAH PROVINSI KALIMANTAN TENGAH**” adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Mei 2018

Allantutra Guslaw  
NRP. 07111650067003

*Halaman ini sengaja dikosongkan*



**PREDIKSI TEMUAN DENGAN METODE TRANSFORMASI  
MASALAH:  
PADA KASUS PEMERIKSAAN LAPORAN KEUANGAN WILAYAH  
PROVINSI KALIMANTAN TENGAH**

Nama mahasiswa : Allantutra Guslaw  
NRP : 07111650067003  
Pembimbing : 1. Dr. Ir. Endroyono, DEA  
2. Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

**ABSTRAK**

Pemeriksaan Laporan Keuangan (LK) oleh Badan Pemeriksa Keuangan Republik Indonesia (BPK RI), salah satunya di wilayah Provinsi Kalimantan Tengah (Kalteng), memiliki beberapa kendala, diantaranya jangka waktu pemeriksaan yang singkat dengan tim pemeriksaan yang terbatas dan data pengelolaan keuangan yang sangat banyak. Sehingga informasi terkait akun yang diperkirakan terdapat temuan-temuan signifikan yang berdampak pada opini dapat menjadi alat bantu (*tool*) bagi pemeriksa dalam menentukan akun LK yang akan disampling untuk diperiksa lebih mendalam. Penelitian sebelumnya terkait pemeriksaan LK masih menggunakan klasifikasi label tunggal, seperti prediksi opini, identifikasi opini, dan deteksi opini. Penelitian ini menggunakan Metode Transformasi Masalah dari klasifikasi multi label. Dengan menggunakan 33 (tiga puluh tiga) atribut berupa rasio keuangan yang diambil dari Laporan Keuangan Pemerintah Daerah (LKPD) di wilayah Provinsi Kalteng Tahun Anggaran (TA) 2006 s.d. 2016, dan 7 (tujuh) label yaitu opini dan temuan berpengaruh opini (pengecualian) yang diambil dari Laporan Hasil Pemeriksaan atas LKPD di wilayah Provinsi Kalteng TA 2007 s.d. 2016. Penelitian ini menggunakan 3 (tiga) pengklasifikasi Metode Transformasi Masalah, yaitu *Binary Relevance* (BR), *Classifier Chain* (CC), dan *Label Combination* (LC), dikombinasikan dengan 4 (empat) pengklasifikasi dasar seperti J48, SMO, *Random Forest* (RF), dan Naive Bayes (NB). Performa pengklasifikasi dievaluasi menggunakan metode validasi silang 10 lipatan (*10-folds cross-validation*) pada 10 (sepuluh) skema dataset yang diperoleh dari 2 (dua) skema atribut yang dinormalisasi dengan 5 (lima) skema normalisasi. Hasil metrik evaluasi pengujian performa pengklasifikasi nilai terbaik yang diperoleh sebagai berikut: *Hamming Loss* adalah 0.194, *One-Error* adalah 0,270, *Rank Loss* adalah 0,166, dan *Average Precision* adalah 0,787. Diketahui bahwa kinerja yang cukup baik dihasilkan oleh penggunaan kombinasi pengklasifikasi BR-RF. Serta penggunaan skema atribut yang diambil dari sektor publik memberikan hasil lebih baik daripada modifikasi dari sektor swasta.

Kata kunci: klasifikasi multi label, metode transformasi masalah, rasio keuangan, laporan keuangan, laporan hasil pemeriksaan

*Halaman ini sengaja dikosongkan*

**PROBLEM TRANSFORMATION METHODS FOR  
PREDICTION OF AUDIT FINDING:  
CASE FOR FINANCIAL STATEMENTS AUDIT IN CENTRAL  
KALIMANTAN PROVINCE**

By : Allantutra Guslaw  
Student Identity Number : 07111650067003  
Pembimbing : 1. Dr. Ir. Endroyono, DEA  
2. Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

**ABSTRACT**

Financial Statements Audit by the Audit Board of the Republic of Indonesia (BPK RI), one of them in the Central Kalimantan Province, has several obstacles, including the short examination period with the limited examination team and a lot of financial management data. So, the information about predicted account contains significant findings that impact on opinion can be a tool for auditors in determining financial statements account to be sampled for more in-depth examination. Previous research related to financial statements audit mostly used single-label classification, such as opinion prediction, opinion identification, and opinion detection. This research use Problem Transformation Methods of multi-label classification. By using 33 (thirty-three) attributes which are financial ratios derived from financial statements in Central Kalimantan Province for the Year 2006 to 2016, and 7 (seven) labels namely opinion and findings impact on opinion (exceptions) taken from financial statements audit reports in Central Kalimantan Province for the Year 2007 to 2016. This research used 3 (three) classifiers of Problem Transformation Methods, namely Binary Relevance (BR), Classifier Chain (CC), and Label Combination (LC), combined with 4 (four) basic classifiers such as J48, SMO, Random Forest (RF), and Naive Bayes (NB). The classification performance is evaluated using 10-folds cross-validation methods on 10 (ten) dataset schemes obtained from 2 (two) attribute schemes normalized with 5 (five) normalization schemes. The evaluation metrics results of classifier performance testing obtained as follows: Hamming Loss is 0.194, One-Error is 0.270, Rank Loss is 0.166, and Average Precision is 0.787. It is known that good performance is generated by the use of a combination of BR-RF classifiers. And the use of attribute schemes taken from the public sector has better results than modifications from the private sector.

Key words: multi-label classification, problem transformation methods, financial ratios, financial statements, audit reports

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

*Alhamdulillahirobbil'alamin*, puji syukur atas segala limpahan nikmat dan karunia Allah SWT. Sehingga atas berkat rahmat Allah SWT penulis akhirnya dapat menyelesaikan tesis ini. Shalawat dan salam untuk Rasulullah Muhammad SAW beserta keluarga, sahabat, dan para pengikutnya yang *istiqomah* hingga akhir zaman. Terima kasih kepada Bapak (alm), Ibu, Mertua, Istri, Anak, Adik, Ipar dan semua pihak, baik yang disebut atau tidak dalam tesis ini, yang telah memberikan dukungan moril maupun materil demi selesainya tesis ini, *Jazakumullah Khoiron Katsiron*.

Ucapan terima kasih dan penghargaan penulis sampaikan kepada Dr. Ir. Endroyono, DEA selaku pembimbing pertama dan Dr. Supeno Mardi Susiki Nugroho, S.T., M.T. selaku pembimbing kedua, yang dengan penuh perhatian, dan kesabaran selalu meluangkan waktu, memberikan pengarahan, dan motivasi serta semangat dalam penulisan tesis ini.

Penulis dapat menyelesaikan tesis ini tidak terlepas dari bantuan, kerjasama dan dukungan dari berbagai pihak. Untuk itu penulis menyampaikan terima kasih kepada:

1. Badan Pemeriksa Keuangan Republik Indonesia dan Kementerian Komunikasi dan Informasi Republik Indonesia yang telah memberikan kesempatan dan beasiswa Program Magister (S2) Bidang Keahlian Telematika Konsentrasi Pengelola TIK Pemerintahan (PeTIK) pada Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya;
2. Prof. Ir. Joni Hermana, M.Sc.Es., Ph.D. selaku Rektor Institut Teknologi Sepuluh Nopember Surabaya;
3. Dr. Tri Arief Sardjono, S.T., M.T. selaku Dekan Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya;
4. Dr. Eng. Ardyono Priyadi, S.T., M.Eng. selaku Kepala Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya;

5. Dr. Ir. Wirawan, DEA selaku Kepala Program Studi Pascasarjana Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya;
6. Dr. Adhi Dharma Wibawa, S.T., M.T. selaku Koordinator Bidang Keahlian Telematika-Pengelola TIK Pemerintahan (PeTIK) pada Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya;
7. Eko Setijadi, S.T., M.T., Ph.D. selaku Dosen Wali Akademik untuk mahasiswa Pengelola TIK Pemerintahan (PeTIK) 2016;
8. Seluruh Dosen dan staf Program Studi Magister (S2) Departemen Teknik Elektro, khususnya Bidang Keahlian Telematika-Pengelola TIK Pemerintahan (PeTIK), atas jasa dan pengabdianya dalam mendidik, membimbing dan mendewasakan kami;
9. Rekan-Rekan PeTIK 2016 serta Telematika 2016 atas semua “hal” yang kita lalui bersama, semoga selalu kompak dan saling mendukung, saling mendoakan baik selama perkuliahan dan seterusnya;
10. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah SWT membalas kebaikan semua pihak yang telah memberi kesempatan, dukungan, doa dan bantuan dalam menyelesaikan tesis ini. Penulis menyadari bahwa tesis ini masih jauh dari sempurna, oleh karena itu saran dan kritik yang membangun sangat diharapkan untuk perbaikan dimasa mendatang. Semoga tesis ini memberikan manfaat yang baik bagi para pembacanya.

Surabaya, Mei 2018

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	iii
PERNYATAAN KEASLIAN TESIS .....	v
ABSTRAK .....	vii
ABSTRACT .....	ix
KATA PENGANTAR .....	xi
DAFTAR ISI .....	xiii
DAFTAR GAMBAR .....	xv
DAFTAR TABEL .....	xvii
DAFTAR NOMENKLATUR .....	xix
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan .....	3
1.4 Batasan Masalah .....	4
1.5 Kontribusi .....	4
1.6 Metodologi Penelitian .....	6
BAB 2 KAJIAN PUSTAKA .....	7
2.1 Penelitian Sebelumnya .....	7
2.1.1 Penelitian terkait Prediksi Opini .....	7
2.1.2 Penelitian terkait Identifikasi Opini .....	11
2.1.3 Penelitian terkait Deteksi Opini .....	15
2.2 Badan Pemeriksa Keuangan .....	16
2.2.1 Pemeriksaan Laporan Keuangan .....	17
2.2.2 Temuan Pemeriksaan Berdampak Opini (Pengecualian) .....	19
2.3 Rasio Keuangan .....	20
2.4 Klasifikasi Multi Label ( <i>Multi-label Classification</i> ) .....	20
2.4.1 Binary Relevance (BR) .....	23
2.4.2 Classifier Chain (CC) .....	24
2.4.3 Label Combination (LC) .....	25
2.4.4 Pengklasifikasi Dasar (Base Classifier) .....	26
2.4.5 MEKA 1.9.1 .....	45
2.4.6 Penggunaan MEKA .....	46
2.4.7 Matrik Evaluasi (Evaluation Metrics) .....	50



2.5	Normalisasi Data .....	53
2.6	Validasi Silang ( <i>Cross Validation</i> ).....	54
BAB 3 METODOLOGI PENELITIAN .....		57
3.1	Alur Penelitian.....	57
3.2	Pengumpulan Data .....	58
3.3	Atribut dan Label.....	58
3.3.1	Atribut.....	58
3.3.2	Label .....	59
3.4	Pembersihan Data.....	61
3.5	Skema Atribut.....	61
3.6	Normalisasi.....	61
3.7	Prediksi dengan Klasifikasi Multi Label .....	62
3.8	Pengukuran Hasil Kinerja .....	64
BAB 4 HASIL DAN PEMBAHASAN .....		65
4.1	Pembersihan Data.....	65
4.2	Stastistik Deskriptif Atribut.....	66
4.3	Skema Dataset .....	67
4.4	Klasifikasi menggunakan MEKA .....	67
4.5	Hasil Matriks Evaluasi .....	68
4.5.1	Hasil Hamming Loss .....	68
4.5.2	Hasil One-Error .....	68
4.5.3	Hasil Rank Loss .....	69
4.5.4	Hasil Average Precision .....	70
4.5.5	Rekapitulasi Hasil Evaluasi .....	70
4.6	Analisa Hasil .....	71
BAB 5 PENUTUP .....		73
5.1	Kesimpulan.....	73
5.2	Saran.....	74
DAFTAR PUSTAKA.....		77
LAMPIRAN .....		81

## DAFTAR GAMBAR

Gambar 1.1 Proses Prediksi Temuan Mendukung Proses Pemeriksaan .....	5
Gambar 2.1 Perbedaan <i>Single-label</i> dengan <i>Multi-label</i> (Herrera et al, 2016).....	21
Gambar 2.2 Pengklasifikasi <i>Binary Relevance</i> (BR) (Herrera et al., 2016) .....	24
Gambar 2.3 Pengklasifikasi <i>Classifier Chain</i> (CC) .....	25
Gambar 2.4 Pengklasifikasi <i>Label Combination</i> (LC) (Herrera et al., 2016).....	26
Gambar 2.5 Data dua kelas dapat dipisahkan secara linier .....	31
Gambar 2.6 Ilustrasi <i>hyperplane</i> dan <i>margin</i> .....	32
Gambar 2.7 Ilustrasi <i>Support Vector</i> (vector pendukung) .....	34
Gambar 2.8 Data dua kelas tidak dapat dipisahkan secara linier .....	36
Gambar 2.9 Pengklasifikasi Multi Label dan dasar pada MEKA v.1.9.1 .....	46
Gambar 2.10 Buka File Dataset dalam format CSV .....	47
Gambar 2.11 Rubah Tipe Data <i>Numeric</i> ke <i>Nominal</i> untuk data Label .....	47
Gambar 2.12 Menentukan Label untuk Klasifikasi .....	48
Gambar 2.13 Simpan dataset dalam format ARFF .....	48
Gambar 2.14 Memilih Classifier yang akan digunakan.....	49
Gambar 2.15 Pilih Metode Evaluasi .....	49
Gambar 2.16 Hasil ditampilkan oleh MEKA.....	50
Gambar 2.17 Contoh Validasi Silang 5 Lipatan (Helix, 2017).....	55
Gambar 3.1 Alur Penelitian.....	57
Gambar 3.2 Contoh opini dan Pengecualian pada LHP LKPD (Buku I).....	60
Gambar 3.3 Pengaturan Tambahan untuk evaluasi .....	63

*Halaman ini sengaja dikosongkan*

## DAFTAR TABEL

Tabel 2.1 Daftar Atribut Prediksi Opini Audit.....	8
Tabel 2.2 <i>Rules</i> (aturan) hasil Prediksi Opini Audit .....	9
Tabel 2.3 Daftar Atribut Prediksi Opini Audit Berkualitas .....	10
Tabel 2.4 Daftar Atribut Identifikasi Opini Audit Berkualitas .....	12
Tabel 2.5 Daftar Atribut Identifikasi Kualifikasi Laporan Audit .....	14
Tabel 2.6 Daftar atribut Deteksi Opini Sektor Publik.....	15
Tabel 2.7 Perbedaan LK Berbasis Kas dan Berbasis Akrua.....	19
Tabel 2.8 Daftar Modifikasi Rasio Keuangan pada sektor pemerintah .....	20
Tabel 3.1 Daftar Atribut dalam penelitian Prediksi Temuan .....	59
Tabel 3.2 Daftar Label untuk penelitian Prediksi Temuan .....	60
Tabel 3.3 Skema Atribut .....	61
Tabel 3.4 Skema Normalisasi setiap dataset .....	62
Tabel 3.5 Pengklasifikasi Metode Transformasi Masalah yang digunakan.....	63
Tabel 4.1 Atribut bernilai tak hingga ( $\infty$ ) yang diganti dengan nilai 0 (nol) .....	65
Tabel 4.2 Statistik Deskriptif Atribut.....	66
Tabel 4.3 Hasil <i>Hamming Loss</i> dari setiap dataset .....	68
Tabel 4.4 Hasil <i>One-Error</i> dari setiap dataset .....	69
Tabel 4.5 Hasil <i>Rank Loss</i> dari setiap dataset.....	69
Tabel 4.6 Hasil <i>Average Precision</i> pada setiap dataset.....	70
Tabel 4.7 Hasil evaluasi untuk setiap skema dataset .....	71
Tabel 4.8 Statistik Deskriptif Label yang diurutkan .....	72

*Halaman ini sengaja dikosongkan*

## DAFTAR NOMENKLATUR

$\Delta$	=	Melambangkan adanya perbedaan
$\Sigma$	=	Penjumlahan dari semua nilai
$\forall_i$	=	<i>Universal quantification</i> , benar untuk semua nilai input
$\wp$	=	Himpunan label ( $L$ ) yang mungkin diperoleh
$\Theta_j$	=	Kumpulan variabel acak dan independen dengan nilai $j = 1, 2, \dots, J$
A1 s.d. A33	=	Atribut yang digunakan, merupakan rasio keuangan yang diperoleh dari nilai Akun-Akun LK
Akun LK	=	Pos-pos yang terdapat di dalam Laporan Keuangan
<i>Algorithm</i>	=	Metode Adaptasi Algoritma
<i>Adaptation Methods</i>		
$AP$	=	<i>Average Precision</i> , menghitung berapa rata-rata banyak posisi yang harus diperiksa sebelum menemukan label yang tidak relevan
APBD	=	Anggaran Pendapatan dan Belanja Daerah
APBN	=	Anggaran Pendapatan dan Belanja Negara
$Argmax$	=	<i>Arguments of the maxima</i> , fungsi yang menentukan nilai input untuk memperoleh hasil nilai output sebesar mungkin (maksimal)
AUC	=	<i>Area Under Curve</i> , menghitung luas area kurva ROC
BPK RI	=	Badan Pemeriksa Keuangan Republik Indonesia
BR	=	<i>Binary Relevance</i>
CaLK	=	Catatas atas Laporan Keuangan
CC	=	<i>Classifier Chain</i>
Entitas	=	Wilayah pemeriksaan

<i>Entropy</i>	=	Menghitung keseragaman suatu sampel
F1	=	Skema atribut yang merupakan modifikasi dari sektor swasta
F2	=	Skema atribut yang diambil dari sektor publik
<i>HL</i>	=	<i>Hamming Loss</i> , menghitung kesalahan prediksi label
Information gain	=	Kriteria untuk pemilihan atribut di algoritma pohon keputusan ( <i>decision tree</i> )
<i>Instance</i>	=	Dataset
Itama	=	Inspektorat Utama
J48	=	Implementasi C4.5 di WEKA/MEKA, sedangkan C4.5 merupakan pengembangan dari ID3, yaitu algoritma untuk membangun sebuah pohon keputusan ( <i>decision tree</i> )
Juklak	=	Petunjuk Pelaksanaan
Juknis	=	Petunjuk Teknis
K/L	=	Kementerian/Lembaga
L1	=	Label yang digunakan merupakan opini, diambil dari LHP LKPD
L2 s.d L7	=	Label yang digunakan merupakan akun LK yang menjadi pengecualian, diambil dari LHP LKPD
LAK	=	Laporan Arus Kas
LC	=	<i>Label Combination</i>
Lebar <i>margin</i>	=	Jarak antara dua kelas data yang dapat dipisahkan oleh garis lurus
LHP	=	Laporan Hasil Pemeriksaan
LHP LKPD	=	Laporan Hasil Pemeriksaan atas Laporan Keuangan Pemerintah Daerah
LK	=	Laporan Keuangan
LKPD	=	Laporan Keuangan Pemerintah Daerah
LO	=	Laporan Operasional



LPE	=	Laporan Perubahan Ekuitas
LPSAL	=	Laporan Perubahan Saldo Anggaran Lebih
LRA	=	Laporan Realisasi Anggaran
MLC	=	Klasifikasi Multi Label ( <i>Multi Label Classification</i> )
MLD	=	Dataset untuk multi label ( <i>Multi Label Dataset</i> )
N1 s.d. N5	=	Skema normalisasi untuk normalisasi nilai atribut
OE	=	<i>One-Error</i> , menghitung nilai positif palsu ( <i>false positif</i> )
OPD	=	Organisasi Perangkat Daerah
Pemda	=	Pemerintah Daerah
Pempus	=	Pemerintah Pusat
<i>Problem Transformation Methods</i>	=	Metode Transformasi Masalah
Rasio Keuangan	=	Perbandingan antara nilai-nilai akun-akun Laporan Keuangan
RL	=	<i>Rank Loss</i> , menghitung kombinasi Label relevan dan tidak relevan setiap <i>instance</i>
ROC Curve	=	<i>Receiver Operating Characteristic Curve</i> , Kurva yang menggambarkan grafik perbandingan antara persentase dari jumlah total positif pada sumbu vertikal terhadap persentase dari jumlah total negatif pada horizontal sumbu.
SDM	=	Sumber Daya Manusia
SMO	=	<i>Sequential Minimal Optimizatio</i> , memecahkan masalah pemrograman kuadratik yang besar menjadi serangkaian masalah pemrograman kuadratik yang kecil
<i>Testing Instance</i>	=	Dataset uji
TMP - <i>Disclaimer</i>	=	Tidak Memberikan Pendapat

<i>Training Instance</i>	=	Dataset latih
TSE	=	<i>Tehran Stock Exchange</i> (Bursa Efek Teheran)
TW - <i>Adversed</i>	=	Tidak Wajar
UU	=	Undang Undang
UUD 1945	=	Undang Undang Dasar Tahun 1945
WDP - <i>Qualified</i>	=	Wajar Dengan Pengecualian
WTP - <i>Unqualified</i>	=	Wajar Tanpa Pengecualian

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Badan Pemeriksa Keuangan Republik Indonesia (BPK RI) memiliki tugas untuk melakukan pemeriksaan atas pengelolaan dan pertanggungjawaban keuangan negara/daerah di Indonesia. Salah satu bentuk atau implementasi pelaksanaan tugas BPK RI adalah kegiatan Pemeriksaan Laporan Keuangan (LK) pada Pemerintahan Pusat, Kementerian/Lembaga dan Pemeriksaan Laporan Keuangan Pemerintah Daerah (LKPD) pada semua Pemerintah Daerah di Indonesia. Pemeriksaan LK/LKPD dilakukan rutin setiap tahun terhadap pengelolaan Anggaran Pendapatan dan Belanja Negara (APBN) oleh Pemerintah Pusat (Pempus) dan Kementerian/Lembaga (K/L) serta terhadap pengelolaan Anggaran Pendapatan dan Belanja Daerah (APBD) oleh Pemerintah Daerah (Pemda) Provinsi/Kabupaten/Kota. Pemeriksaan LK/LKPD dilakukan untuk memberikan opini yaitu penilaian pemeriksa atas kewajaran informasi keuangan yang disajikan pada LK/LKPD tersebut atas pengelolaan dan pertanggungjawaban keuangan oleh Pempus/K/L/Pemda.

Pemeriksaan LK/LKPD dilakukan untuk mengidentifikasi masalah dalam pengelolaan dan pertanggungjawaban keuangan negara/daerah yang akan diungkap sebagai temuan-temuan di Laporan Hasil Pemeriksaan atas Laporan Keuangan (LHP LK/LKPD). Dari masalah-masalah yang teridentifikasi selama proses pemeriksaan atau temuan-temuan yang diungkap dalam LHP LK/LKPD akan dapat mempengaruhi pemberian/penentuan opini atas penyajian LK/LKPD oleh Pempus/K/L/Pemda tersebut, sebagai hasil pelaksanaan kegiatan Pemeriksaan Laporan Keuangan. Opini untuk LK/LKPD secara umum terdiri atas Wajar Tanpa Pengecualian (WTP), Wajar Dengan Pengecualian (WDP), Tidak Wajar (TW) dan Tidak Memberikan Pendapat (TMP). Untuk opini selain WTP juga akan disebutkan akun LK yang menjadi pengecualian berdasarkan temuan-temuan signifikan/berpengaruh yang diungkap di LHP LK/LKPD. Untuk opini WTP, tidak disebutkan akun LK yang menjadi pengecualian, karena tidak ada akun yang menjadi pengecualian dalam pemberian opini WTP tersebut.

Untuk membantu pelaksanaan tugasnya dalam memeriksa pengelolaan dan pertanggungjawaban keuangan negara/daerah, BPK RI memiliki kantor perwakilan di setiap Provinsi di Indonesia, salah satunya adalah di Provinsi Kalimantan Tengah. BPK Perwakilan Provinsi Kalimantan Tengah memiliki tanggung jawab untuk melakukan pemeriksaan atas pengelolaan dan pertanggungjawaban keuangan daerah pada 15 (lima belas) entitas Pemda Provinsi/Kabupaten/Kota di wilayah Provinsi Kalimantan Tengah. Kesulitan yang dihadapi dalam pelaksanaan proses pemeriksaan adalah adanya keterbatasan waktu, Sumber Daya Manusia (SDM) dengan pendidikan dan pengalaman pemeriksaan yang berbeda dan banyak dokumen keuangan yang harus diperiksa. Serta susunan tim pemeriksa yang ditugaskan untuk melaksanakan kegiatan pemeriksaan LKPD di sebuah entitas (Pemda) selalu berubah setiap tahunnya.

Pemeriksaan LK/LKPD dilakukan 2 (dua) tahap, yaitu pemeriksaan pendahuluan dan terinci. Pemeriksaan pendahuluan dilakukan sebelum Pempus/K/L/Pemda menyelesaikan penyusunan Laporan Keuangan. Sedangkan pemeriksaan terinci dilaksanakan setelah Pempus/K/L/Pemda menyelesaikan penyusunan LK/LKPD dan menyerahkan LK/LKPD kepada BPK RI.

Pemeriksaan pendahuluan dan terinci masing-masing dilakukan dalam jangka waktu 30 (tiga puluh) s.d. 40 (empat puluh) hari kalender. Sedangkan susunan tim pemeriksaan terdiri dari 1 (satu) Ketua Tim dan antara 2 (dua) s.d. 5 (lima) anggota Tim, tergantung dari jumlah pegawai/pemeriksa yang ada di BPK Perwakilan Provinsi Kalimantan Tengah yang bisa diikutkan dalam pelaksanaan tugas pemeriksaan. Dokumen pengelolaan dan pertanggungjawaban keuangan yang diperiksa terdiri dari dokumen yang dimiliki banyak Organisasi Perangkat Daerah (OPD) dan unit kerja – unit kerja di bawahnya pada Pemda yang ada di wilayah Provinsi Kalimantan Tengah.

Pegawai/pemeriksa yang ditugaskan dalam tim pemeriksaan terdiri dari bermacam latar belakang pendidikan dengan berbagai pengalaman pemeriksaan yang berbeda. Untuk pegawai/pemeriksa yang sudah sering/terbiasa mengikuti tugas pemeriksaan dan memiliki intuisi pemeriksaan yang bagus, akan dapat memperkirakan akun LK mana yang harus disampling untuk diperiksa lebih mendalam karena diprediksi akan teridentifikasi masalah (terdapat temuan) yang

signifikan/berpengaruh pada pemberian opini. Dan tidak semua pegawai/pemeriksa yang ikut dalam tim pemeriksaan memiliki kemampuan yang baik dalam memperkirakan akun LK yang harus disampling.

Dengan adanya penelitian Prediksi Temuan ini diharapkan mampu menghasilkan alat (*tool*) yang dapat membantu tim pemeriksa dalam menentukan akun LK yang akan disampling/diperiksa karena diprediksi/diperkirakan teridentifikasi masalah/temuan yang signifikan/berpengaruh pada pemberian opini sehingga harus diperiksa secara lebih.

Penelitian sebelumnya terkait Pemeriksaan LK/LKPD hanya memberikan informasi berupa perkiraan/prediksi opini yang akan diperoleh/diberikan atas penyajian LK/LKPD. Penelitian tersebut diantaranya adalah prediksi opini, identifikasi opini, dan deteksi opini yang menggunakan metode klasifikasi label tunggal (*single label classification*). Hasil prediksi opini tersebut belum mampu memberikan informasi yang dapat membantu dalam menentukan akun LK yang akan disampling untuk diperiksa. Dalam penelitian Prediksi Temuan ini menggunakan Metode Transformasi Masalah yang merupakan salah satu pendekatan dari klasifikasi multi label (*multi label classification*).

## **1.2 Rumusan Masalah**

Belum tersedianya informasi berupa akun LK yang diperkirakan terdapat temuan-temuan signifikan yang berdampak pada opini, sehingga dapat digunakan sebagai alat bantu dalam Pemeriksaan LK. Informasi tersebut dapat membantu dalam menentukan akun LK yang menjadi prioritas untuk disampling/diperiksa lebih mendalam karena diperkirakan/diprediksi terdapat masalah/temuan yang signifikan sehingga proses pemeriksaan menjadi lebih terarah.

## **1.3 Tujuan**

Tujuan penelitian ini adalah untuk memberikan informasi berupa hasil prediksi yaitu akun LK yang diperkirakan akan menjadi pengecualian karena terdapat temuan-temuan signifikan yang berdampak/berpengaruh terhadap opini

dengan menggunakan Metode Transformasi Masalah, yang merupakan salah satu pendekatan Klasifikasi Multi Label. Sehingga dapat digunakan sebagai alat bantu (*tool*) bagi pemeriksa dalam menentukan akun LK yang akan disampling pada pelaksanaan tugas Pemeriksaan LK.

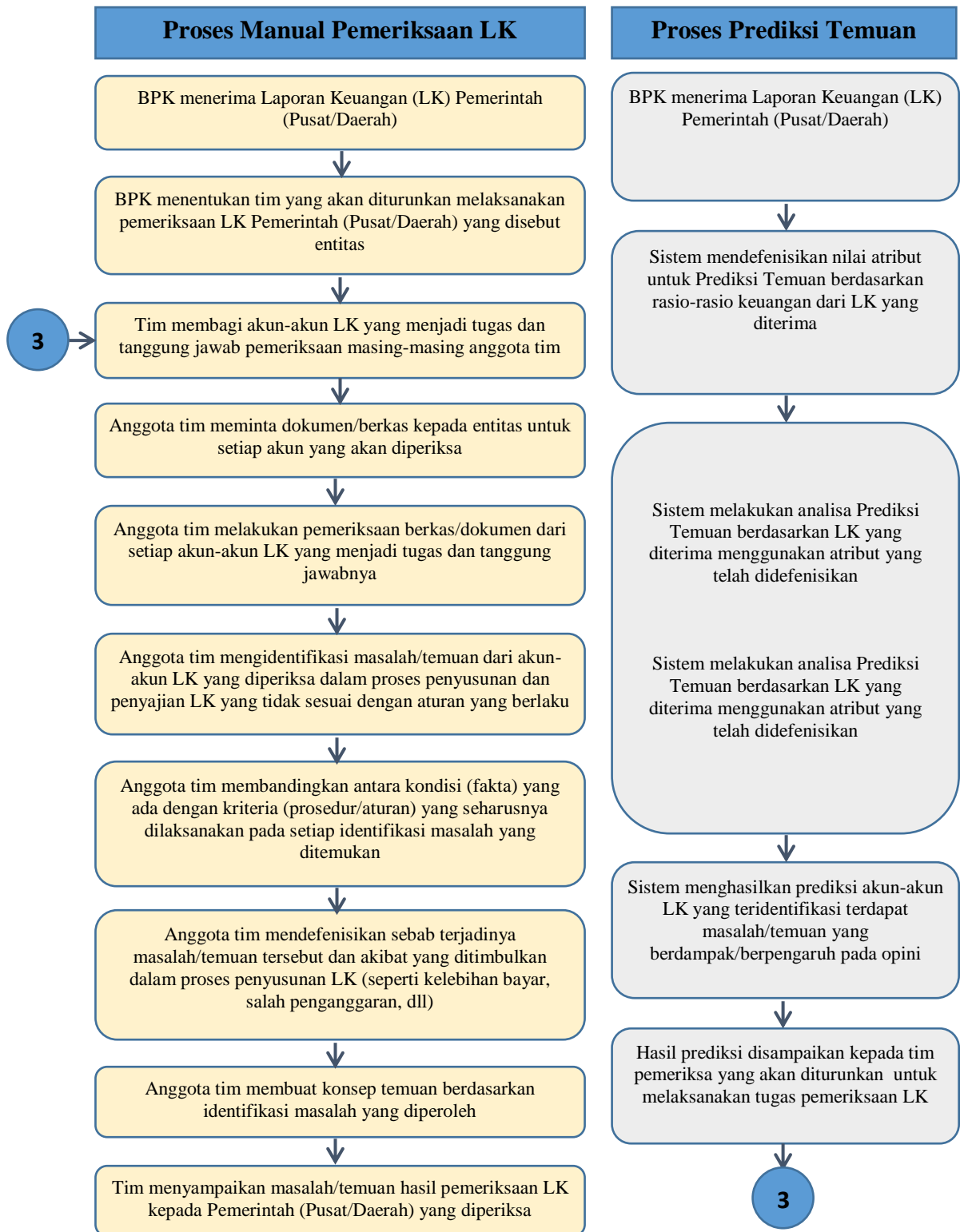
#### **1.4 Batasan Masalah**

Penelitian ini menggunakan 33 (tiga puluh tiga) atribut berupa rasio keuangan yang diperoleh dari data Laporan Keuangan Pemerintah Daerah (LKPD) di wilayah Provinsi Kalimantan Tengah Tahun Anggaran 2006 s.d. 2016, dan 7 (tujuh) label yaitu opini dan temuan berpengaruh opini (pengecualian) yang diambil dari data LHP LKPD Tahun Anggaran 2007 s.d. 2016 di wilayah Provinsi Kalimantan.

#### **1.5 Kontribusi**

Penelitian ini diharapkan dapat memberikan kontribusi dalam menyediakan alat (*tool*) yang mampu memberikan informasi awal, berupa hasil prediksi akun LK yang diperkirakan akan menjadi pengecualian karena terdapat temuan-temuan signifikan yang berdampak/ berpengaruh pada opini guna membantu pemeriksa dalam menentukan akun LK yang akan disampling/diperiksa lebih mendalam dalam tugas pemeriksaan LK. Dengan harapan proses Pemeriksaan LK dapat menjadi lebih terarah, sehingga waktu pemeriksaan yang terbatas dapat dimanfaatkan dengan lebih baik. Juga diharapkan dengan adanya penelitian ini dapat menjadi salah satu sarana dalam memanfaatkan data pertanggungjawaban dan pengelolaan keuangan negara/daerah yang telah dikumpulkan menggunakan aplikasi e-Audit guna menghasilkan informasi-informasi bermanfaat yang mampu mendukung tugas pokok BPK RI dalam melakukan pemeriksaan atas pengelolaan dan pertanggungjawaban keuangan negara/daerah.

Alur proses Prediksi Temuan dalam mendukung proses Pemeriksaan diilustrasikan pada Gambar 1.1.



Gambar 1.1 Proses Prediksi Temuan Mendukung Proses Pemeriksaan



## 1.6 Metodologi Penelitian

Penelitian ini menggunakan Metode Transformasi Masalah yang merupakan salah satu pendekatan Klasifikasi Multi Label untuk melakukan prediksi akun LK yang diperkirakan akan menjadi pengecualian karena terdapat temuan-temuan signifikan yang berdampak/berpengaruh terhadap opini. Atribut pada dataset berupa rasio keuangan yang diambil dari LKPD, sedangkan label adalah opini dan pengecualian yang dinyatakan pada LHP LKPD dari 15 (lima belas) Pemda di wilayah Provinsi Kalimantan Tengah selama 10 tahun, yaitu tahun 2007 s.d. 2016. Dataset yang diperoleh berupa atribut dan label dibagi menjadi 2 (dua) skema atribut.

Setiap skema atribut akan dinormalisasi dengan 5 (lima) skema normalisasi, sehingga diperoleh 10 skema dataset. Pengklasifikasi Metode Transformasi Masalah yang digunakan adalah *Binary Relevance* (BR), *Classifier Chain* (CC), dan *Label Combination* (LC), yang masing-masing akan dikombinasikan dengan 4 (empat) pengklasifikasi dasar seperti J48, SMO, *Random Forest* (RF), dan *Naive Bayes* (NB). Maka diketahui terdapat 12 (dua belas) kombinasi pengklasifikasi yang digunakan. Untuk mengukur kinerja/performa setiap kombinasi pengklasifikasi yang diujikan pada setiap skema dataset, nilai metrik evaluasi yang dihitung adalah *Hamming Loss*, *One-Error*, *Rank Loss*, dan *Average Precision*. Rincian metode dan langkah penelitian akan dipaparkan pada Bab 3.

## **BAB 2**

### **KAJIAN PUSTAKA**

#### **2.1 Penelitian Sebelumnya**

Penelitian sebelumnya terkait Pemeriksaan Laporan Keuangan sebagian besar masih menggunakan metode klasifikasi label tunggal (*single label classification*), diantaranya adalah penelitian tentang prediksi opini (*opinions prediction*), identifikasi opini (*identifying audit opinions*) dan deteksi opini (*opinion detection*). Penelitian-penelitian tersebut juga menggunakan rasio keuangan yang diperoleh dari rasio/perbandingan nilai-nilai pada akun-akun Laporan Keuangan (LK) sebagai atribut, tetapi masih menggunakan 1 (satu) label yaitu opini pemeriksaan. Berikut adalah beberapa jurnal terkait penelitian sebelumnya yang menjadi acuan penulis dalam menentukan atribut yang digunakan dalam melakukan penelitian Prediksi Temuan ini.

##### **2.1.1 Penelitian terkait Prediksi Opini**

###### **2.1.1.1 Prediksi Opini Audit**

Penelitian prediksi opini audit (Saif et al., 2012) adalah melakukan prediksi opini hasil pemeriksaan dengan menggunakan data Laporan Keuangan beberapa perusahaan di Iran yang diperoleh dari *Tehran Stock Exchange* (TSE). Penelitian tersebut pada awalnya mengambil data hasil pemeriksaan terhadap perusahaan-perusahaan yang terdaftar di *Tehran Stock Exchange*, yaitu 1018 hasil pemeriksaan Laporan Keuangan dari Tahun 2001 s.d 2007 yang terdiri atas 708 data dengan opini Wajar Dengan Pengecualian (WDP/*qualified*) dan 310 data dengan opini Wajar Tanpa Pengecualian (WTP/*unqualified*). Namun setelah mengecualikan perusahaan keuangan seperti Bank, perusahaan asuransi, lalu dipilih perusahaan dalam daftar yang minimal pernah mendapatkan opini Wajar Dengan Pengecualian (WDP/*qualified*) antara Tahun 2001 s.d. 2007, dan menutup pembukuan pada Maret 2007. Data yang diperoleh untuk penelitian tersebut dibagi menjadi dua set yaitu data latih (*training*) dan data uji (*test*).

Atribut yang digunakan dalam melakukan penelitian Prediksi Opini Audit dapat dilihat pada Tabel 2.1.

Tabel 2.1 Daftar Atribut Prediksi Opini Audit

Atribut	Uraian
A1	Z-Score*
A2	Log net sales
A3	Log total assets
A4	Log number of employees
A5	Current ratio
A6	Quick ratio
A7	Total debts to total assets
A8	Working capital per employee
A9	Total assets per employee
A10	Net sales per employee
A11	Profit per employee
A12	Debtors turnover
A13	Debtor collection period
A14	Net assets turnover
A15	Fixed assets turnover
A16	EBIT margin
A17	Earnings before tax margin
A18	Cash from operating activities to net sales
A19	Cash from investing activities to net sales
A20	Return on equity
A21	Equity to long-term debts
A22	Return on total assets
A23	Return on capital employed
A24	Inventories turnover
A25	Tax payables to sales
A26	Provision for staff termination benefits per employee
A27	Retained earnings to net sales
A28	Litigation
A29	Growth
$*Z\text{-Score} = 3.20784K1 + 1.80384K2 + 1.61363K3 + 0.50094K4 + 0.16903K5 - 0.39709K6 - 0.12505K7 + 0.33849K8 + 1.42363K9$	
K1	EBIT /Total assets
K2	Retained earnings /Total assets
K3	Working capital /Total assets
K4	Equity /Total debts
K5	EBIT /Net sales
K6	Current assets /Current liabilities
K7	Net profit /Net sales
K8	Total debts /Total assets
K9	Company size

Sumber: Saif et al., 2012

Penelitian tersebut bertujuan untuk menemukan *rules*/aturan dari sebuah mesin cerdas yang mampu membantu pemeriksa dalam memprediksi opini hasil pemeriksaan menggunakan data Laporan Keuangan. Dengan menggunakan algoritma baru yang merupakan gabungan dari metode *Support Vector Machine* (SVM) dengan metode *Decision Tree* (DT). Algoritma SVM digunakan sebagai

*preprocesor* untuk membantu *Decision Tree* dalam memilih data (*instance*) agar menghasilkan ekstraksi aturan (*rules*) yang akan digunakan untuk prediksi opini. 30 (tiga puluh) aturan sebagai hasil dari penelitian tersebut dalam melakukan prediksi opini hasil pemeriksaan dapat dilihat pada Tabel 2.2.

Tabel 2.2 *Rules* (aturan) hasil Prediksi Opini Audit

No.	Kondisi (IF)	Hasil (THEN)
1.	A11 < 54,48; A4 < 7,00; A29 < 1,17; A13 < 24,29	Qualified (WDP)
2.	A11 < 54,48; A4 < 7,00; A29 < 1,17; A13 > 24,29; A2 < 9,48	
3.	A11 < 54,48; A4 < 7,00; A29 < 1,17; A13 > 24,29; A2 > 9,48; A10 < 69,37	
4.	A11 < 54,48; A4 < 7,00; A29 < 1,17; A13 > 24,29; A2 > 9,48; A10 > 69,37; A14 < -5,89	
5.	A11 < 54,48; A4 < 7,00; A29 > 1,17	
6.	A11 < 54,48; A4 > 7,00; A6 < 0,48; A21 < 0,61; A1 > 17,49	
7.	A11 < 54,48; A4 > 7,00; A6 < 0,48; A5 > 0,63	
8.	A11 > 54,48; A12 < 2,46; A10 < 363,94	
9.	A11 > 54,48; A12 < 2,46; A10 > 363,94; A11 < 58,43	
10.	A11 > 54,48; A12 > 2,46; A24 < 0,10	
11.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 < -0,44	
12.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 < 0,15; A26 > 35,48	
13.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 > 0,15; A1 < 18,58; A22 < 0,24; A1 < 18,28	
14.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 > 0,15; A1 > 18,58	
15.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 > 13,39; A3 < 14,73	
16.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A27 > 1,62	
17.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A21 < 0,46; A14 > 1,67; A1 > 18,47	
18.	A11 < 54,48; A4 < 7,00; A29 < 1,17; A13 > 24,29; A2 > 9,48; A10 > 69,37; A14 > -5,89	UnQualified (WTP)
19.	A11 < 54,48; A4 > 7,00; A6 < 0,48; A21 < 0,61	
20.	A11 < 54,48; A4 > 7,00; A6 < 0,48; A21 < 0,61; A1 < 17,49	
21.	A11 < 54,48; A4 > 7,00; A6 < 0,48; A5 < 1,63	
22.	A11 > 54,48; A12 < 2,46; A10 > 363,94; A11 > 58,43	
23.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 < 0,15; A26 < 35,48	
24.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 > 0,15; A1 < 18,58; A22 < 0,24; A1 > 18,28	
25.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 < 13,39; A22 > 0,15; A1 < 18,58; A22 > 0,24	
26.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A2 > 13,39; A3 > 14,73	
27.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A8 < 201,92; A27 < 1,62	
28.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A21 < 0,46; A14 < 1,67	
29.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A21 < 0,46; A14 > 1,67; A1 < 18,47	
30.	A11 > 54,48; A12 > 2,46; A24 > 0,10; A19 > -0,44; A12 < 5,81; A21 > 0,46	

Sumber: Saif et al., 2012

### 2.1.1.2 Prediksi Opini Audit Berkualitas

Penelitian prediksi opini hasil pemeriksaan berkualitas/Wajar Dengan Pengecualian (WDP/*qualified*) (Yaşar et al., 2015) menggunakan data rasio keuangan pada beberapa perusahaan di Turki yang diperoleh dari *Istanbul Stock Exchange* (ISE). Data yang digunakan adalah 110 Laporan Keuangan dari Tahun 2010 s.d 2013 terdiri atas 55 data dengan opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*) dan 55 data dengan opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*). Daftar atribut yang digunakan dalam melakukan penelitian tersebut dapat dilihat pada Tabel 2.3.

Tabel 2.3 Daftar Atribut Prediksi Opini Audit Berkualitas

Atribut	Uraian
A1	Current assets to current liabilities
A2	Cash from operating activities to total assets
A3	Working capital to total assets
A4	Receivables plus inventories to total assets
A5	Total liabilities to total assets
A6	Equity to total liabilities
A7	Net sales to total assets
A8	Net sales to equity
A9	Net income to total assets
A10	Retained earnings to total assets
A11	Earnings before interest and taxes to total assets
A12	Net income to equity

Sumber: Yaşar et al., 2015

Penelitian tersebut bertujuan untuk memprediksi opini hasil pemeriksaan WDP/*qualified* menggunakan metode *discriminant*, *logistic regression (logit)* dan C5.0 *decision tree*, dengan hasil sebagai berikut:

1. Hasil analisa univariat menunjukkan bahwa perusahaan yang memperoleh opini hasil pemeriksaan WDP/*qualified* memiliki rasio likuiditas lebih rendah, ekuitas yang rendah terhadap total kewajiban, produktivitas operasi yang rendah dan kontribusi yang lebih rendah.
2. Hasil analisa menggunakan metode *discriminant* menunjukkan bahwa variabel A10 adalah variabel yang paling efisien untuk memisahkan antara data perusahaan yang memperoleh opini hasil pemeriksaan WTP/*unqualified* dengan WDP/*qualified*, dan variabel yang signifikan lainnya adalah A6, A12, A7, dan A3. Model *discriminant* mengklasifikasikan dengan benar 92,7% opini hasil pemeriksaan

WTP/*unqualified* dan 81,8% opini hasil pemeriksaan WDP/*qualified*, dengan keberhasilan klasifikasi total model adalah 87,3%.

3. Hasil pemodelan dengan *logistic regression* menunjukkan variabel A10, A6, dan A3 secara statistik merupakan variabel yang signifikan untuk memprediksi opini hasil pemeriksaan WDP/*qualified*. Dari hasil analisa dapat diindikasikan bahwa perusahaan dengan tingkat likuiditas, profitabilitas dan *leverage* keuangan yang lebih tinggi, kemungkinan besar akan menurunkan probabilitas opini hasil pemeriksaan WDP/*qualified*. Keakuratan hasil prediksi adalah 98,2% untuk opini hasil pemeriksaan WTP/*unqualified* dan 87,3% untuk opini hasil pemeriksaan WDP/*qualified*, dan total keberhasilan klasifikasi model sebesar 92,7%.
4. Hasil analisa menggunakan metode C5.0 *decision tree* mengindikasikan bahwa variabel A10 adalah yang paling efisien dalam memprediksi opini hasil pemeriksaan. Variabel penting lainnya yang berpengaruh besar dalam menentukan opini hasil pemeriksaan adalah A5, A12, A9. Metode C5.0 *decision tree* dapat mengklasifikasikan secara benar 96,4% opini hasil pemeriksaan WTP/*unqualified* dan 100% opini hasil pemeriksaan WDP/*qualified* sehingga keberhasilan klasifikasi total model adalah 98,2%.

## **2.1.2 Penelitian terkait Identifikasi Opini**

### **2.1.2.1 Identifikasi Opini Audit Berkualitas**

Penelitian tersebut melakukan identifikasi opini audit yang berkualitas (WDP) (Pourheydari et al., 2012) menggunakan 4 (empat) metode klasifikasi penambangan data untuk menemukan model yang bisa mengidentifikasi opini hasil pemeriksaan dengan menggunakan data keuangan pada beberapa perusahaan yang diperoleh dari *Tehran Stock Exchange* (TSE) dari rentang waktu tahun 2000 s.d. 2007. Data yang digunakan sebanyak 1018 Laporan Keuangan yang terdiri atas 347 data dengan opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*) dan 671 data dengan opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*). Untuk memperoleh validasi model yang baik, data *sample* dibagi menjadi dua set data, yaitu data *training* yang terdiri atas 814

data Laporan Keuangan dan data *test* terdiri atas 204 data Laporan Keuangan. Daftar atribut yang digunakan dalam melakukan penelitian tersebut ditampilkan pada Tabel 2.4.

Tabel 2.4 Daftar Atribut Identifikasi Opini Audit Berkualitas

Atribut	Uraian
A1	Z-Score*
A2	Log net sales
A3	Log total assets
A4	Log number of employees
A5	Current ratio
A6	Quick ratio
A7	Solvency ratio
A8	Working capital per employee
A9	Total assets per employee
A10	Net sales per employee
A11	Profit per employee
A12	Account receivable turnover
A13	Account receivable collection period
A14	Net assets turnover
A15	Fixed assets turnover
A16	EBIT margin
A17	EBT margin
A18	Cash flows from operating to sales
A19	Cash flow from investment activities to sales
A20	Return on shareholders' funds
A21	Ratio of equity to sum of long term debts and current share of long term debts
A22	Return on total assets
A23	Return on capital employed
A24	Inventories turnover
A25	Tax Allowance to sale ratio
A26	Post retirement benefits allowance per employee
A27	Retained earnings to net sales
A28	Firm Litigation
A29	Firm's Growth

Sumber: Pourheydari et al., 2012

Penelitian tersebut bertujuan memfasilitasi pemeriksa agar mampu melakukan identifikasi opini hasil pemeriksaan Laporan Keuangan, namun juga dapat dimanfaatkan oleh komisi bursa efek, sebagai pengawas pasar modal, untuk mengevaluasi kualitas Laporan Keuangan perusahaan-perusahaan yang terdaftar di pasar modal. Penelitian tersebut dalam melakukan pengujian menggunakan 4 (empat) metode yaitu *multi-layer perceptron neural network* (MLP), *probabilistic neural network* (PNN), *radial basic function network* (RBF) dan *logistic regression* (LR), dengan hasil sebagai berikut:



1. MLP memiliki kemampuan yang tinggi dalam mengidentifikasi dan memprediksi jenis opini hasil pemeriksaan, dengan tingkat akurasi 87,75%.
2. PNN adalah model yang paling seimbang dalam mengidentifikasi jenis opini hasil pemeriksaan, dan memiliki sedikit kesalahan dalam mengidentifikasi jenis opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*) dan Wajar Dengan Pengecualian (WDP/*qualified*), dibandingkan dengan model lainnya.
3. RBF memiliki presisi prediksi 85,30%, dibandingkan dengan model lainnya, merupakan kinerja tertinggi dalam mengidentifikasi jenis opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*), namun model lain memiliki kinerja yang lebih baik dalam mengidentifikasi opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*).
4. LR memiliki kinerja terlemah dalam mengidentifikasi opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*), dan merupakan model yang tidak seimbang dalam mengidentifikasi jenis opini hasil pemeriksaan yang berbeda.

#### 2.1.2.2 Identifikasi Kualifikasi Laporan Audit

Penelitian identifikasi kualifikasi laporan auditor (Zdolšek et al., 2015) dilakukan untuk mengkaji hasil pemeriksaan atas Laporan Keuangan dalam kaitannya dengan keadaan dalam situasi ekonomi perusahaan yang mengarah pada kualifikasi Laporan Keuangan (opini hasil pemeriksaan). Data yang digunakan adalah hasil pemeriksaan atas Laporan Keuangan Tahun 2009 yang diterbitkan pada Tahun 2010 dari 293 perusahaan, yang terdiri atas 12 data dengan opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*) dan 281 data dengan opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*). Data yang digunakan tersebut diperoleh dari *database* perusahaan-perusahaan di Slovenia, serta merupakan perusahaan besar yang memiliki kantor pusat di Slovenia

Daftar atribut yang digunakan dalam melakukan penelitian tersebut dapat dilihat pada Tabel 2.5.

Tabel 2.5 Daftar Atribut Identifikasi Kualifikasi Laporan Audit

Atribut	Uraian
A1	Debt ratio
A2	Debt to equity ratio
A3	Long-term financing to long-term assets ratio
A4	Quick ratio
A5	Liquidity ratio
A6	Current ratio
A7	Short-term business receivables to short-term business liabilities ratio
A8	Inventory turnover ratio
A9	Trade receivables turnover ratio
A10	Cash turnover
A11	Operating efficiency ratio
A12	Total assets turnover
A13	Working capital turnover
A14	Current liabilities turnover
A15	Capital employed turnover
A16	Short-term assets rate
A17	Cash assets rate
A18	Return on assets rate
A19	Total assets
A20	Auditor's report time lag (in days)

Sumber: Zdolšek et al., 2015

Penelitian tersebut bertujuan untuk mengidentifikasi opini hasil pemeriksaan Laporan Keuangan perusahaan di Slovenia guna menunjukkan keadaan/situasi ekonomi perusahaan yang dapat berpengaruh pada opini dengan menggunakan model *logistic regression (logit)* yang dikembangkan dari rasio keuangan/akuntansi perusahaan. Hasilnya menunjukkan bahwa perusahaan dengan opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*) memiliki hutang yang tinggi, likuiditas rendah, efisiensi rendah dan profitabilitas yang buruk dibandingkan dengan perusahaan dengan opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*). Dari sudut pandang statistik, model *logistic regression* dapat membedakan antara perusahaan yang menerima opini hasil pemeriksaan Wajar Dengan Pengecualian (WDP/*qualified*) dengan yang menerima opini hasil pemeriksaan Wajar Tanpa Pengecualian (WTP/*unqualified*) dari data sampel perusahaan besar di Slovenia.

### 2.1.3 Penelitian terkait Deteksi Opini

#### 2.1.3.1 Deteksi Opini Sektor Publik

Penelitian Deteksi Opini Sektor Publik (Arianto et al., 2007a) atau Klasifikasi Data Keuangan Sektor Publik (Arianto, 2007b) terkait penentuan sampel oleh Inspektorat Utama (Itama) BPK RI dalam pemeriksaan internal atas proses pemeriksaan LK. Data yang digunakan adalah LHP LK TA 2007 s.d. 2014 pada 150 entitas yang dipilih secara acak, terdiri atas 75 data dengan opini WTP, 25 data dengan opini WDP, 25 data dengan opini TW, dan 25 data dengan opini TMP. Daftar atribut yang digunakan disajikan pada Tabel 2.6.

Tabel 2.6 Daftar atribut Deteksi Opini Sektor Publik

Atribut	Uraian
A1	Hutang
A2	Total Aset
A3	SILPA (Sisa Lebih Pembiayaan Anggaran)
A4	PAD (pendapatan asli daerah)
A5	Kas Tunai dan Kas Bank
A6	Piutang
A7	Kas Tunai dan Kas Bank /PAD
A8	Kas Tunai dan Kas Bank /Total Aset
A9	SILPA /Total Aset
A10	Aset Lancar /Total Aset
A11	SILPA /PAD
A12	Piutang /PAD
A13	PAD /Total Aset
A14	Aset Lancar /Hutang Lancar
A15	PAD /Aset Tetap
A16	Kas tunai /Total Aset
A17	Kas Tunai dan Kas Bank /Hutang Lancar
A18	Total Hutang /Total Ekuitas
A19	Hutang Jangka Panjang /Total Aset
A20	Total Hutang /Total Aset
A21	Total Aset /Total Ekuitas
A22	Hutang Jangka Panjang /Total Ekuitas
A23	Aset Tetap /Total Aset
A24	Kas Tunai & Kas Bank /Aset Lancar
A25	Piutang /Total Aset
A26	Total Belanja /Total Belanja Tahun Lalu
A27	PAD /PAD Tahun Lalu
A28	Piutang /Piutang Tahun Lalu
A29	Total Aset /Total Aset Tahun Lalu
A30	Belanja Modal /Perubahan Nilai Aset Tetap
A31	Belanja Barang Jasa /Perubahan Nilai Persediaan
A32	Belanja Pegawai /Total Belanja
A33	Belanja Modal /Total Belanja
A34	Belanja Hibah /Total Belanja
A35	Belanja Bantuan Sosial /Total Belanja
A36	PAD /Pendapatan Transfer
A37	Zona Wilayah
A38	Jenis Wilayah Administratif

Sumber: Arianto, 2007b

Penelitian tersebut bertujuan untuk membandingkan penggunaan atribut modifikasi dari sektor swasta (A1 s.d. A29) dan atribut murni dari sektor pemerintah (A30 s.d. A38). Penelitian deteksi opini hasil pemeriksaan ini menggunakan metode *k-Nearest Neighbor (kNN)*. Dataset yang diperoleh dibagi menjadi 9 (sembilan) skema atribut (fitur) yang kemudian masing-masingnya dinormalisasi dengan 5 (lima) skema normalisasi. Juga membandingkan hasil klasifikasi 2 (dua) kelas (Label: WTP dan WDP) dan 4 (empat) kelas (Label: WTP, WDP, TW, dan TMP). Hasilnya penelitian tersebut diantaranya sebagai berikut:

1. Metode klasifikasi yang dihasilkan baik untuk 2 kelas maupun 4 kelas tergolong buruk karena memiliki nilai *Area Under Curve (AUC)* tertinggi pada rentang 0,60 s.d. 0,70.
2. Hasil AUC tertinggi untuk klasifikasi 2 kelas sebesar 66,30% dengan penggunaan kombinasi normalisasi data menggunakan eksponensial ZScore, jarak euclidean, fitur/atribut A2, A4, A6, A7, A11, A13, A15, A24, A26, A35, A36 dengan nilai k terbaik adalah k=2.
3. Hasil uji statistik untuk 2 kelas, maka skema terbaik adalah kombinasi normalisasi menggunakan Z-Score, jarak euclidean, fitur/atribut A13, A15, A36, dan nilai k=1.
4. Hasil AUC tertinggi untuk 4 kelas sebesar 61,56% dengan penggunaan kombinasi normalisasi data dengan skop (0 s.d. 1), jarak euclidean, fitur/atribut A30 s.d. A38 dan nilai k terbaik adalah k=2.
5. Hasil uji statistik untuk 4 kelas, maka skema terbaik adalah kombinasi normalisasi menggunakan eksponensial Z-Score, jarak euclidean, fitur/atribut A30 s.d. A38 dan nilai k terbaik adalah k=2.
6. Penggunaan atribut murni dari sektor pemerintahan memiliki hasil lebih baik dibandingkan atribut modifikasi dari sektro swasta.

## **2.2 Badan Pemeriksa Keuangan**

Badan Pemeriksa Keuangan Republik Indonesia (BPK RI) memiliki tugas dalam pemeriksaan terhadap pengelolaan keuangan negara tercantum dalam Undang Undang Dasar Negara Republik Indonesia Tahun 1945 (UUD 1945)

(Perubahan Ketiga) Pasal 23 E ayat (1) “Untuk memeriksa pengelolaan dan tanggung jawab tentang keuangan negara diadakan satu Badan Pemeriksa Keuangan yang bebas dan mandiri”, sedangkan pada UUD 1945 sebelum perubahan tercantum dalam pasal 23 ayat (5) “Untuk memeriksa tanggung jawab tentang keuangan negara diadakan suatu Badan Pemeriksa Keuangan, yang peraturannya ditetapkan dengan undang-undang. Hasil pemeriksaan itu diberitahukan kepada Dewan Perwakilan Rakyat”.

Tugas dan wewenang BPK RI dituangkan dalam UUD 1945, Perubahan Ketiga) (Bab VIII A) dengan tiga pasal (23E, 23F, dan 23G) juga didukung dengan Undang Undang (UU) di bidang Keuangan Negara yaitu UU No.1 Tahun 2004 Tentang Perbendaharaan Negara, UU Nomor 17 Tahun 2003 tentang Keuangan Negara, UU No.15 Tahun 2004 Tentang Pemeriksaan Pengelolaan dan Tanggung Jawab Keuangan Negara, dan Nomor 15 Tahun 2006 Tentang Badan Pemeriksa Keuangan serta Peraturan-Peraturan, Petunjuk Pelaksanaan (Juklak) dan Petunjuk Teknis (Juknis) yang diterbitkan oleh BPK RI sendiri.

Berdasarkan UUD 1945, BPK RI memiliki kantor perwakilan di setiap provinsi di Indonesia, salah satunya adalah di Provinsi Kalimantan Tengah, tempat penulis bekerja sebelum melaksanakan tugas belajar di Institut Teknologi Sepuluh Nopember Surabaya. BPK Perwakilan Provinsi Kalimantan Tengah memiliki 15 (lima belas) entitas pemeriksaan yaitu Provinsi Kalimantan Tengah, Kota Palangka Raya, Kabupaten Kapuas, Kabupaten Pulang Pisau, Kabupaten Gunung Mas, Kabupaten Barito Selatan, Kabupaten Barito Timur, Kabupaten Barito Utara, Kabupaten Murung Raya, Kabupaten Katingan, Kabupaten Kotawaringin Timur, Kabupaten Kotawaringin Barat, Kabupaten Seruyan, Kabupaten Lamandau, dan Kabupaten Sukamara.

### **2.2.1 Pemeriksaan Laporan Keuangan**

Tugas pemeriksaan yang dilakukan oleh BPK RI terbagi kepada 3 (tiga) jenis, yaitu Pemeriksaan Laporan Keuangan, Pemeriksaan Kinerja dan Pemeriksaan Dengan Tujuan Tertentu. Pada penelitian ini penulis melakukan prediksi akun Laporan Keuangan (LK) yang diperkirakan terdapat temuan-temuan signifikan yang berdampak pada opini, karena temuan-temuan signifikan yang

merupakan hasil dari identifikasi masalah pada kegiatan Pemeriksaan Laporan Keuangan dapat mempengaruhi pemberian opini.

Pemeriksaan Laporan Keuangan adalah kegiatan pemeriksaan rutin yang dilakukan setiap tahun oleh BPK RI secara khusus oleh BPK Perwakilan Provinsi Kalimantan Tengah pada 15 (lima belas) entitas (Pemerintah Provinsi/Kabupaten/Kota) di wilayah Provinsi Kalimantan Tengah. Hasil pemeriksaan Laporan Keuangan tersebut adalah opini atas penyajian Laporan Keuangan oleh Pemerintah Daerah (Pemda). Laporan Hasil Pemeriksaan atas Laporan Keuangan Pemerintah Daerah (LHP LKPD) sebagai hasil kegiatan pemeriksaan Laporan Keuangan (LK) terdiri atas 3 (tiga) buku. Buku I berisi opini penyajian LKPD dan LK Audited (Laporan Keuangan yang sudah diperiksa/di-audit). Buku II adalah hasil pemeriksaan terhadap Sistem Pengendalian Intern (SPI) yang berisi temuan-temuan terkait SPI berdasarkan penyajian Laporan Keuangan di Pemda tersebut. Buku III adalah hasil pemeriksaan atas Kepatuhan Terhadap Peraturan Perundang-undangan berdasarkan penyajian Laporan Keuangan yang dilaksanakan Pemda tersebut. Pada Buku I LHP LKPD tidak dicantumkan temuan pemeriksaan karena berisi opini dan pengecualian atas penyajian Laporan Keuangan oleh Pemda dan LK Audited (Laporan Keuangan yang sudah diperiksa/di-audit). Temuan pemeriksaan LK dituliskan pada Buku II LHP LKPD terkait SPI dan Buku III LHP LKPD terkait Kepatuhan Terhadap Peraturan Perundang-undangan.

Penyajian Laporan Keuangan Pemerintah Daerah pada Tahun Anggaran 2006 s.d. 2014 adalah berbasis Kas (*cash basis*), sedangkan semenjak Tahun Anggaran 2015 penyajian LKPD adalah berbasis akrual (*accrual basis*). Komponen pada Laporan Keuangan berbasis Kas diantaranya adalah Neraca, Laporan Realisasi Anggaran (LRA), Laporan Arus Kas (LAK) dan Catatan atas Laporan Keuangan (CaLK). Sedangkan komponen pada Laporan Keuangan berbasis akrual diantaranya adalah Laporan Realisasi Anggaran (LRA), Laporan Perubahan Saldo Anggaran Lebih (LPSAL), Neraca, Laporan Operasional (LO), Laporan Arus Kas (LAK), Laporan Perubahan Ekuitas (LPE), dan Catatan atas Laporan Keuangan (CaLK).

Perbedaan komponen yang disajikan pada Laporan Keuangan (LK) berbasis Kas dan AkruaI dapat dilihat pada Tabel 2.7.

Tabel 2.7 Perbedaan LK Berbasis Kas dan Berbasis AkruaI

No.	Uraian	Berbasis Kas	Berbasis AkruaI
1.	Neraca	✓	✓
2.	Laporan Realisasi Anggaran (LRA)	✓	✓
3.	Laporan Perubahan Saldo Anggaran Lebih (LPSAL)	-	✓
4.	Laporan Operasional (LO)	-	✓
5.	Laporan Arus Kas (LAK)	✓	✓
6.	Laporan Perubahan Ekuitas (LPE)	-	✓
7.	Catatan atas Laporan Keuangan (CaLK)	✓	✓

### 2.2.2 Temuan Pemeriksaan Berdampak Opini (Pengecualian)

Seperti yang telah dijelaskan sebelumnya bahwa pada pemeriksaan LKPD menghasilkan LHP yang terdiri dari Buku I, II dan III. Pada Buku I LHP LKPD berisi opini dan pengecualian penyajian Laporan Keuangan oleh Pemda, dan Laporan Keuangan Pemda yang telah diperiksa/di-*audit*. Pada Buku II dan III berisi temuan-temuan yang diperoleh pada saat pelaksanaan pemeriksaan terkait SPI dan Kepatuhan Terhadap Peraturan Perundang-undangan dalam penyajian Laporan Keuangan oleh Pemda tersebut.

Setiap temuan yang dituliskan pada setiap LHP memiliki format yang sama, yaitu terdapat penjelasan terkait kondisi, kriteria, akibat, sebab, tanggapan dan rekomendasi. Kondisi menggambarkan keadaan yang terjadi (fakta) yang berkaitan dengan temuan yang disampaikan. Kriteria memuat peraturan dan undang-undang yang berkaitan dengan keadaan yang diceritakan pada kondisi. Akibat menjelaskan hal-hal yang terjadi akibat pelanggaran terhadap kriteria tersebut. Sebab menjelaskan penyebab terjadinya kondisi tersebut. Tanggapan berisi tanggapan dari Pemda (OPD terkait) atas permasalahan yang terjadi. Rekomendasi berisi hal-hal yang harus dilakukan oleh Pemda untuk menindaklanjuti hasil temuan tersebut.

Pada Buku I LHP LKPD dijelaskan terkait opini dan pengecualian yang merupakan akun-akun LK yang terdapat temuan-temuan signifikan yang berdampak/berpengaruh terhadap penentuan opini penyajian LK tersebut. Penelitian Prediksi Temuan ini opini dan pengecualian yang merupakan akun-

akun LK yang terdapat temuan-temuan signifikan yang berdampak/berpengaruh terhadap penentuan opini tersebut dijadikan label dalam dataset yang akan diuji.

### 2.3 Rasio Keuangan

Dalam penelitian ini, atribut yang digunakan merupakan rasio keuangan yang diperoleh dari nilai akun-akun yang terdapat dalam Laporan Keuangan. Rasio keuangan tersebut adalah perbandingan antara nilai-nilai akun-akun pada Laporan Keuangan. Contoh atribut yang akan digunakan dapat dilihat pada Tabel 2.1, Tabel 2.3, Tabel 2.4, dan Tabel 2.5. Daftar akun pada Laporan Keuangan Pemerintah Daerah (LKPD) yang dapat dijadikan acuan dalam penentuan atribut untuk penelitian ini dapat dilihat pada Lampiran 1, Lampiran 2, dan Lampiran 3. Beberapa rasio keuangan sektor swasta yang dapat dimodifikasi ke sektor pemerintahan dapat dilihat pada Tabel 2.8.

Tabel 2.8 Daftar Modifikasi Rasio Keuangan pada sektor pemerintah

No.	Atribut	Uraian	Modifikasi Sektor Pemerintahan
1.	Current Ratio	Current Assets / CurrentLiability	Jumlah Aset Lancar /Jumlah Kewajiban Jangka Pendek
2.	Quick Ratio	(Current assets – Inventories) /Current liabilities	(Jumlah Aset Lancar – Persediaan) /Jumlah Kewajiban Jangka Pendek
3.	Debt Ratio	Total liabilities /Total assets	Jumlah Kewajiban /Jumlah Aset
4.	Working Capital	Current Assets – Currnet Liabilities	Jumlah Aset Lancar – Jumlah Kewajiban Jangka Pendek
5.	Working Capital Turnover	Sales /Net Working Capital	Jumlah Pendapatan /(Jumlah Aset Lancar – Jumlah Kewajiban Jangka Pendek)
6.	Inventory turnover	Sales /Inventories	Jumlah Pendapatan /Persediaan
7.	Fixed assets turnover	Sales /Net fixed assets	Jumlah Pendapatan /Jumlah Aset Tetap
8.	Total assets turnover	Sales /Total assets	Jumlah Pendapatan /Jumlah Aset
9.	Return on equity (ROE)	Net income /equity	SILPA /Jumlah Ekuitas Dana
10.	Return on total assets (ROA)	Net income available /Total assets	SILPA /Jumlah Aset
11.	Cash Turnover	Net Sales /Cash	Jumlah Pendapatan /Kas dan Setara Kas
12.	Accounts Receivable Turnover	Net Sales /Average Gross Receivables	Jumlah Pendapatan /Piutang

Sumber: Ehrhardt & Brigham, 2011

### 2.4 Klasifikasi Multi Label (*Multi-label Classification*)

Klasifikasi Multi Label (Abdallah et al., 2015; Kavitha &Mahalekshmi, 2016) berbeda dengan klasifikasi label tunggal. Pada Klasifikasi Multi Label



setiap dataset (*instance*) atau kumpulan atribut pada dataset terkait dengan beberapa label yang terdiri dari beberapa kelas biner. Sedangkan dalam klasifikasi label tunggal setiap dataset (*instance*) atau kumpulan atribut dalam dataset hanya terkait dengan satu label, baik itu kelas biner (*binary-class*) maupun multi kelas (*multi-class*). Perbedaan antara klasifikasi label tunggal (*single label classification*) dengan klasifikasi multi label (*multi label classification*) diilustrasikan pada Gambar 2.1.

Dataset Kelas Biner						Dataset Multi Kelas					
$X_1$	$X_2$	...	$X_{f-1}$	$X_f$	$Y$	$X_1$	$X_2$	...	$X_{f-1}$	$X_f$	Class
$X_{11}$	$X_{12}$	...	$X_{1f-1}$	$X_{1f}$	0	$X_{11}$	$X_{12}$	...	$X_{1f-1}$	$X_{1f}$	C1
$X_{21}$	$X_{22}$	...	$X_{2f-1}$	$X_{2f}$	1	$X_{21}$	$X_{22}$	...	$X_{2f-1}$	$X_{2f}$	C2
...					0	...					C3
$X_{a1}$	$X_{a2}$	...	$X_{af-1}$	$X_{af}$	1	$X_{a1}$	$X_{a2}$	...	$X_{af-1}$	$X_{af}$	C4

Dataset Multi Label							
$X_1$	$X_2$	...	$X_{f-1}$	$X_f$	$Y_1$	...	$Y_k$
$X_{11}$	$X_{12}$	...	$X_{1f-1}$	$X_{1f}$	1	0	1
$X_{21}$	$X_{22}$	...	$X_{2f-1}$	$X_{2f}$	0	1	1
...					1	0	0
$X_{a1}$	$X_{a2}$	...	$X_{af-1}$	$X_{af}$	0	1	0

Gambar 2.1 Perbedaan *Single-label* dengan *Multi-label* (Herrera et al, 2016)

Klasifikasi multi label (Tsoumakas & Katakis, 2007) pada awalnya dimotivasi oleh tugas untuk melakukan kategorisasi teks dan diagnosis medis. Karena dokumen teks dapat memiliki lebih dari satu kelas konseptual. Contoh, artikel surat kabar mengenai reaksi gereja Kristen terhadap rilis film Da Vinci Code dapat diklasifikasikan ke dalam kategori *Society\Religion and Arts\Movies*. Demikian pula dalam diagnosis medis, seorang pasien mungkin menderita misalnya diabetes dan kanker prostat pada saat bersamaan. Metode klasifikasi multi label saat ini semakin dibutuhkan oleh aplikasi modern, seperti untuk klasifikasi fungsi protein, kategorisasi musik dan klasifikasi semantik foto (Boutella et al., 2004). Dalam klasifikasi semantik foto, sebuah foto dapat memiliki lebih dari satu kelas konseptual, seperti matahari terbenam dan pantai pada saat yang sama. Demikian pula dalam pengkategorian musik, sebuah lagu dapat memiliki lebih dari satu genre. Sebagai contoh, beberapa lagu hit dari band *rock* populer Scorpions dapat dicirikan sebagai *rock* dan *balada*.

Terdapat beberapa asumsi yang berkaitan dengan masalah multi label (Read, 2010; Gibaja & Ventura, 2010) diantaranya adalah sebagai berikut:

1. Kumpulan label sudah ditentukan sebelumnya, sehingga label tersebut memiliki arti dan dapat ditafsirkan manusia (meskipun bukan oleh seorang ahli), dan semua label juga relevan dengan domain masalah.
2. Jumlah label mungkin terbatas dalam ruang lingkup, sedemikian rupa, sehingga dapat dijelajahi oleh manusia, dan biasanya tidak lebih banyak dari jumlah atribut.
3. Setiap contoh/data pelatihan (*training instance*) dapat dikaitkan dengan sejumlah label dari set label yang ditentukan.
4. Jumlah atribut yang mewakili contoh pelatihan dapat bervariasi, tetapi tidak perlu mempertimbangkan kasus ekstrim dari ribuan atribut karena strategi pengurangan atribut dapat digunakan dalam kasus-kasus ini.
5. Jumlah contoh pelatihan mungkin besar, karena pengklasifikasi multi label mungkin harus menghadapi ratusan ribu contoh yang potensial.
6. Korelasi antara label dapat dimungkinkan.
7. Data mungkin tidak seimbang.

Pada Klasifikasi Multi Label terdapat 2 (dua) pendekatan utama (Abdallah et al., 2015; Kavitha & Mahalekshmi, 2016), sebagai berikut:

1. Metode Transformasi Masalah (*Problem Transformation Methods*), yaitu mentransformasi masalah multi label menjadi beberapa (satu atau lebih) masalah label tunggal.

Contoh Metode Transformasi Masalah (*Problem Transformation Methods*) antara lain:

- *Binary Relevance (BR)*,
- *Classifier Chain (CC)* (Read et al., 2009),
- *Classifier Trellis (CT)* (Read et al., 2015),
- *Label Combination (LC)*,
- *Calibrated Label Ranking (CLR)*.

2. Metode Adaptasi Algoritma (*Algorithm Adaptation Method*), yaitu mengadaptasi algoritma label tunggal dan membuat penyesuaian agar dapat menghasilkan keluaran multi label.

Contoh Metode Adaptasi Algoritma (*Algorithm Adaptation Method*) antara lain:

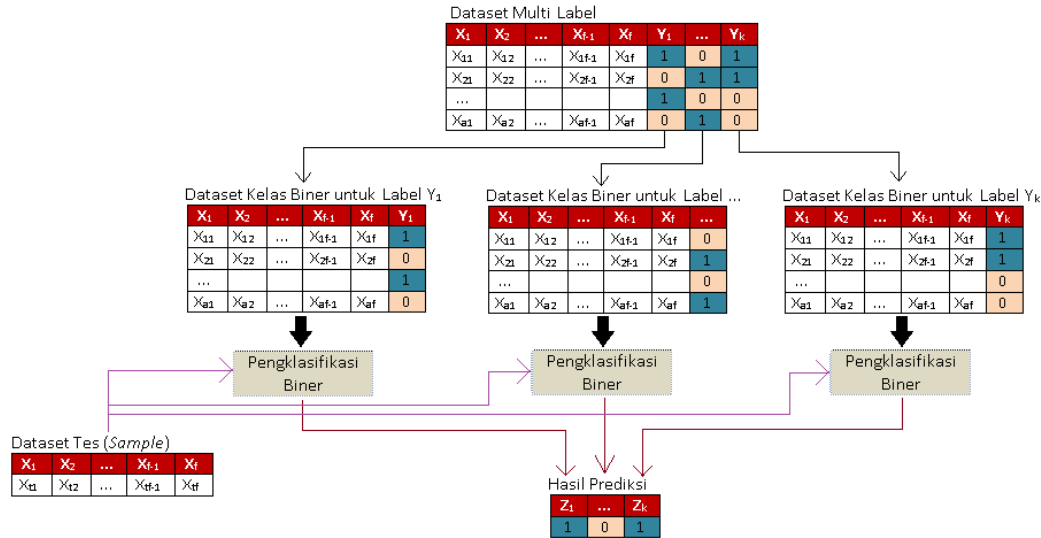
- *Multi-label k-Nearest Neighbor* (ML-kNN) (Zhang & Zhou, 2007),
- *Multi-label Decision Tree* (ML-DT),
- *Ranking Support Vector Mesin* (Rank-SVM),
- *Collective Multi-label* (CML).

#### **2.4.1 Binary Relevance (BR)**

*Binary Relevance* (BR) merupakan Metode Transformasi Masalah yang paling umum. *Binary Relevance* mempelajari pengklasifikasi biner (*biner classifier*), dengan menguraikan masalah multi label menjadi beberapa masalah label tunggal kelas biner (*binary-class*) yang dibagi beberapa label. Setiap pengklasifikasi biner (*classifier biner*) kemudian bertanggung jawab untuk memprediksi menggunakan asosiasi label tunggal.

Pada dasarnya BR memperlakukan setiap label sebagai masalah klasifikasi label tunggal yang terpisah. Metode ini adalah yang paling sederhana dan efisien, tetapi satu-satunya kelemahan dari metode ini adalah tidak mempertimbangkan korelasi antar label karena diperlakukan secara independen. Dataset multi label awal dibagi menjadi beberapa dataset kelas biner sebanyak label yang ada, lalu digunakan untuk melatih pengklasifikasi biner (label tunggal). Untuk setiap sampel uji baru, akan diberikan kepada setiap pengklasifikasi biner yang telah dilatih. Lalu hasil setiap prediksi mereka digabung untuk mendapatkan hasil klasifikasi label akhir.

Diagram transformasi menggunakan pengklasifikasi *Binary Relevance* (BR) dapat diilustrasikan seperti pada Gambar 2.2.

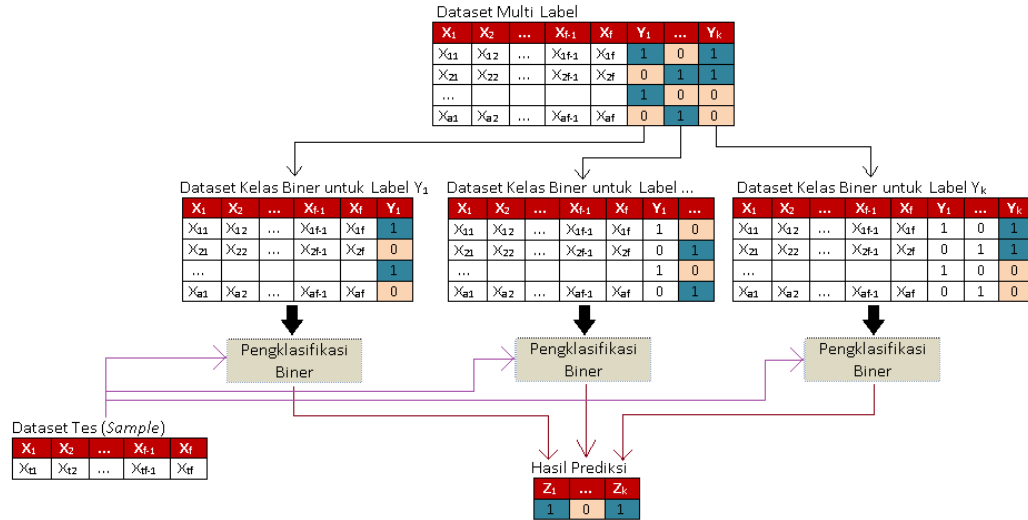


Gambar 2.2 Pengklasifikasi *Binary Relevance* (BR) (Herrera et al., 2016)

## 2.4.2 Classifier Chain (CC)

Metode *Classifier Chain* (CC) mengubah masalah multi-label menjadi beberapa masalah klasifikasi tunggal dengan *binary-class* yang berantai. *Classifier Chain* (CC) mempertahankan efisiensi komputasi dari metode *Binary Relevance* (BR) dan juga mempertimbangkan ketergantungan label untuk klasifikasi. *Classifier Chain* (CC) melibatkan pengklasifikasi biner sejumlah label yang terhubung secara berantai, dan masing-masing pengklasifikasi menangani masalah relevansi biner yang terkait dengan setiap label. Ruang fitur dari setiap tautan dalam rantai diperluas dengan asosiasi label 0/1 (biner) dari semua tautan sebelumnya. Klasifikasi pertama dilatih hanya pada data input dan kemudian setiap pengklasifikasi (*classifier*) berikutnya dilatih pada ruang input dan semua pengklasifikasi sebelumnya dalam rantai. *Classifier Chain* (CC) sangat mirip dengan *Binary Relevance* (BR), satu-satunya perbedaan adalah ia membentuk rantai untuk mempertahankan korelasi antar label.

Diagram transformasi menggunakan pengklasifikasi *Classifier Chain* (CC) dapat diilustrasikan seperti pada Gambar 2.3.



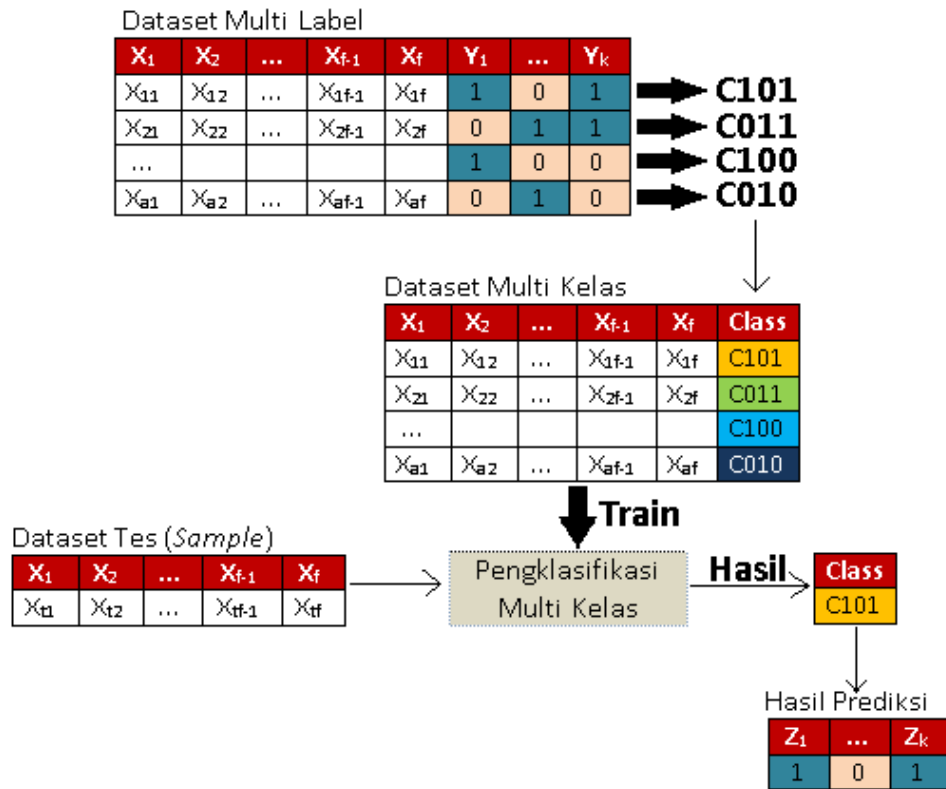
Gambar 2.3 Pengklasifikasi *Classifier Chain* (CC)

#### 2.4.3 Label Combination (LC)

Metode *Label Powerset* (LP) yang juga disebut dengan *Label Combination* (LC) merupakan metode kombinasi label, yang mengubah masalah multi-label menjadi satu masalah label tunggal dengan multi-kelas. Dalam LC masalah multi-label diubah menjadi satu masalah label tunggal multi-class dengan menggunakan nilai target untuk atribut *class* pada semua subset unik dari multi label yang ada dalam *training instance* (subset berbeda dari label). Metode ini mempertimbangkan ketergantungan label yang mempertimbangkan setiap kejadian unik dari set label dalam multilabel dataset *training* sebagai satu kelas untuk dataset yang baru.

Sebagai contoh, jika sebuah *instance* berkaitan dengan tiga label L1, L2, L4, maka kelas label tunggal baru menjadi L1,2,4. Pada hasil transformasi dataset baru berupa masalah klasifikasi label tunggal multi kelas, setiap pengklasifikasi label tunggal dapat diterapkan. Untuk mengklasifikasikan contoh baru, LC menghasilkan kelas yang paling mungkin, yang sebenarnya adalah sekumpulan label. Dengan menganggap ketergantungan label dan tidak ada informasi yang hilang selama klasifikasi. Jika pengklasifikasi dapat menghasilkan distribusi probabilitas atas semua kelas, maka LC dapat memberikan peringkat di antara

semua label yang menggunakan pendekatan ini. Diagram transformasi menggunakan pengklasifikasi *Classifier Chain* (CC) dapat diilustrasikan seperti pada Gambar 2.4.



Gambar 2.4 Pengklasifikasi *Label Combination* (LC) (Herrera et al., 2016)

#### 2.4.4 Pengklasifikasi Dasar (Base Classifier)

Setiap pengklasifikasi multi label yang digunakan pada Metode Transformasi Masalah akan dikombinasikan dengan pengklasifikasi dasar (pengklasifikasi label tunggal) dalam melakukan pengujian pada data multi label. Berikut beberapa contoh pengklasifikasi dasar yang dapat digunakan:

##### 1. J48

Pengklasifikasi J48 (Cherman et al., 2011) adalah membangkitkan pengklasifikasi pohon keputusan (*Decision Tree* - DT) yang memiliki fleksibilitas untuk memilih subset fitur yang berbeda pada simpul internal yang berbeda dari pohon sedemikian rupa sehingga subset fitur dipilih secara optimal membedakan antar kelas dalam simpul menurut beberapa

kriteria. DT memiliki fungsi seleksi fitur yang tertanam (*embedded*), yaitu hanya sebagian fitur yang dipertimbangkan menjadi simpul dalam pohon keputusan. Jika terdapat fitur-fitur tambahan dari contoh-contoh tak berlabel ada pada pohon keputusan, maka pohon keputusan dinilai gagal dalam menemukan ketergantungan pada label.

*Decision Tree* (DT) (Sayad, 2010-2018a) membangun model klasifikasi atau model regresi dalam bentuk struktur pohon yang dapat digunakan untuk tipe data kategori dan numerik.. DT membagi kumpulan data menjadi himpunan bagian yang lebih kecil sementara pada saat yang sama mengembangkan sebuah pohon keputusan secara bertahap. DT pada akhirnya akan menghasilkan sebuah struktur pohon dengan simpul keputusan (*decision node*) dan simpul daun (*leaf node*). Simpul keputusan dapat memiliki dua atau lebih cabang simpul. Simpul daun akan mewakili hasil klasifikasi atau keputusan. Simpul keputusan teratas/pertama di struktur pohon adalah prediktor terbaik yang disebut akar simpul (*root node*).

Algoritma inti untuk membangun pengklasifikasi *Decision Tree* (DT) disebut ID3 oleh J. R. Quinlan yang menggunakan pencarian serampangan dari atas ke bawah (*greedy top-down search*) melalui ruang dari cabang-cabang yang mungkin tanpa melakukan pengecekan mundur (*backtracking*). ID3 menggunakan *entropy* dan informasi *Gain* untuk membangun sebuah pohon keputusan, yang mencakup semua prediktor dengan asumsi adanya ketergantungan antara semua prediktor.

DT dibangun dari atas ke bawah dari sebuah simpul akar yang melibatkan pembagian data ke dalam subset yang berisi data (*instance*) dengan nilai yang sama (*homogen*). Algoritma ID3 menggunakan *entropy* untuk menghitung keseragaman suatu sampel. Jika sampel benar-benar homogen, maka *entropy* akan bernilai 0 (nol) dan jika sampel dapat dibagi sama maka *entropy akan* bernilai 1 (satu). Pohon keputusan dapat diubah menjadi seperangkat aturan dengan memetakan data dari simpul akar ke simpul daun satu demi satu. Untuk membangun pohon keputusan, kita perlu menghitung *entropy* dan *gain* dari setiap atribut yang ada.

Terdapat dua jenis perhitungan *entropy* menggunakan tabel frekuensi sebagai berikut:

- a. Perhitungan *Entropy* dengan tabel frekuensi satu atribut sesuai persamaan (2.1):

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2.1)$$

Keterangan:

- $E$  : *Entropy*  
 $S$  : himpunan kasus  
 $c$  : jumlah partisi/bagian dalam himpunan kasus  
 $p_i$  : proporsi partisi kasus terhadap kasus

- b. Perhitungan *Entropy* dengan tabel frekuensi dua atribut sesuai persamaan (2.2):

$$E(T,X) = \sum_{c \in X} P(c) E(c) \quad (2.2)$$

Informasi *gain* didasarkan pada penurunan *entropy* setelah dataset dibagi pada sebuah atribut. Dalam membangun sebuah pohon keputusan penting untuk menemukan atribut yang memperoleh nilai informasi *gain* tertinggi.

Langkah dalam membangun sebuah pohon keputusan:

1. Hitung *entropy* target.
2. Kumpulan data kemudian dibagi pada atribut yang berbeda. *Entropy* untuk setiap cabang dihitung, lalu ditambahkan secara proporsional, untuk menemukan total *entropy* untuk pembagian. *Entropy* yang dihasilkan dikurangi dari *entropy* sebelum pembagian. Hasilnya adalah informasi *gain*, atau pengurangan *entropy*.

Untuk menghitung nilai informasi *gain* sesuai persamaan (2.3):

$$Gain(S,A) = E(S) - \sum_{i=1}^c \frac{|S_i|}{|S|} * E(S_i) \quad (2.3)$$

Keterangan:

- $S$  : himpunan kasus  
 $A$  : atribut  
 $|S_i|$  : jumlah kasus pada partisi ke- $i$   
 $|S|$  : jumlah kasus dalam  $S$



3. Memilih atribut yang memiliki nilai informasi *gain* terbesar sebagai simpul keputusan, lalu membagi dataset berdasarkan cabangnya dan ulangi proses yang sama untuk setiap cabang.
4. Cabang dengan nilai *entropy* 0 akan menjadi simpul daun.
5. Cabang dengan nilai *entropy* lebih dari 0, akan dilakukan proses pembagian data secara lebih lanjut.
6. Algoritma ID3 dijalankan secara rekursif pada cabang-cabang yang bukan merupakan simpul daun, dan diulang sampai semua data dapat diklasifikasikan.

## 2. SMO

*Sequential Minimal Optimization* (SMO) (Cherman et al., 2011) adalah algoritma mesin vektor pendukung (SVM - *Support Vector Machine*). Algoritma ini dikelola untuk membangun batas keputusan (pemisah margin maksimum) dengan kemungkinan jarak terbesar ke titik contoh. Batas keputusan semacam ini untuk membantu dalam melakukan pemisahan/pengklasifikasian dengan baik.

SVM (Sayad, 2010-2018b) melakukan klasifikasi dengan menemukan *hyperplane* yang memaksimalkan *margin*/jarak antara kedua kelas. Vektor (kasus) yang menentukan *hyperplane* adalah vektor pendukung (*support vector*).

Algoritma SVM adalah sebagai berikut:

- a. Menentukan *hyperplane* optimal dengan memaksimalkan lebar *margin* ( $w$ ). Setelahnya akan dapat menentukan nilai  $w$  dan  $b$  dengan menyelesaikan fungsi obyektif menggunakan Pemrograman Kuadrat.
- b. Memperluas definisi di atas untuk masalah yang tidak dapat dipisahkan secara linier atau memiliki *penalty* untuk kesalahan klasifikasi.
- c. Memetakan data ke ruang dimensi tinggi agar lebih mudah untuk mengklasifikasikan dengan permukaan keputusan linier yaitu merumuskan masalah sehingga data dipetakan secara implisit.

Keindahan SVM adalah bahwa jika data terpisah secara linier, maka akan ada nilai minimum global yang unik. Analisis SVM yang ideal adalah dengan menghasilkan *hyperplane* yang benar-benar dapat memisahkan vektor (kasus) menjadi dua kelas yang tidak tumpang tindih. Bagaimanapun, pemisahan sempurna mungkin tidak dapat dilakukan, atau mungkin menghasilkan model dengan banyak kasus sehingga tidak dapat melakukan klasifikasi dengan benar. Dalam situasi ini, SVM menemukan *hyperplane* yang memaksimalkan *margin* dan meminimalkan kesalahan klasifikasi. Algoritma SVM mencoba mempertahankan variabel *slack* ke nol dengan memaksimalkan *margin*. Namun, itu tidak meminimalkan jumlah *misclassifications* tetapi jumlah jarak dari *margin hyperplanes*.

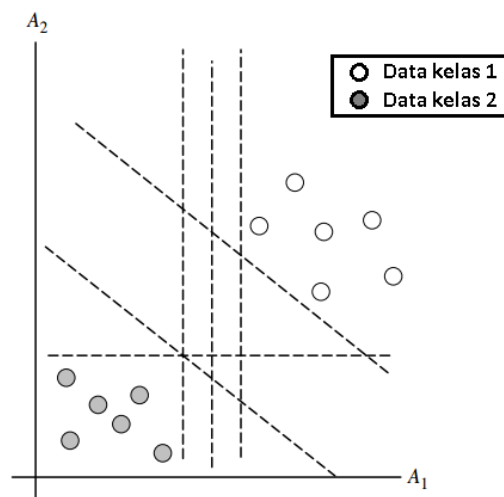
*Support Vector Machines* merupakan metode yang menjanjikan untuk klasifikasi data linier dan nonlinier (Han & Kamber, 2006). Algoritma SVM bekerja dengan menggunakan pemetaan nonlinier untuk mengubah data *training* menjadi dimensi yang lebih tinggi guna pencarian *hyperplane* optimal untuk pemisah secara linier yang memisahkan data satu kelas dengan data kelas yang lain. Dengan pemetaan nonlinier yang sesuai ke dimensi yang cukup tinggi, maka data dari dua kelas akan dapat dipisahkan oleh *hyperplane*. SVM menemukan *hyperplane* dengan menggunakan vektor dukungan dari data *training* dan lebar *margin* yang didefinisikan oleh vektor dukungan.

Landasan awal penggunaan algoritma SVM telah ada dimulai tahun 1960-an, diantaranya oleh Vapnik dan Alexei Chervonenkis tentang teori pembelajaran statistik. Sedangkan penelitian awal tentang SVM dipresentasikan pada tahun 1992 oleh Vladimir Vapnik, Bernhard Boser dan Isabelle Guyon serta lainnya. Waktu pelatihan SVM bisa sangat lambat, tetapi dapat dikatakan akurat, karena adanya kemampuan untuk memodelkan batas keputusan nonlinier yang kompleks.

#### A. SVM untuk data yang dapat dipisahkan secara linier

Contoh penerapan algoritma SVM yang sederhana adalah pada masalah dua kelas yang dapat dipisahkan secara linier. Data yang dapat

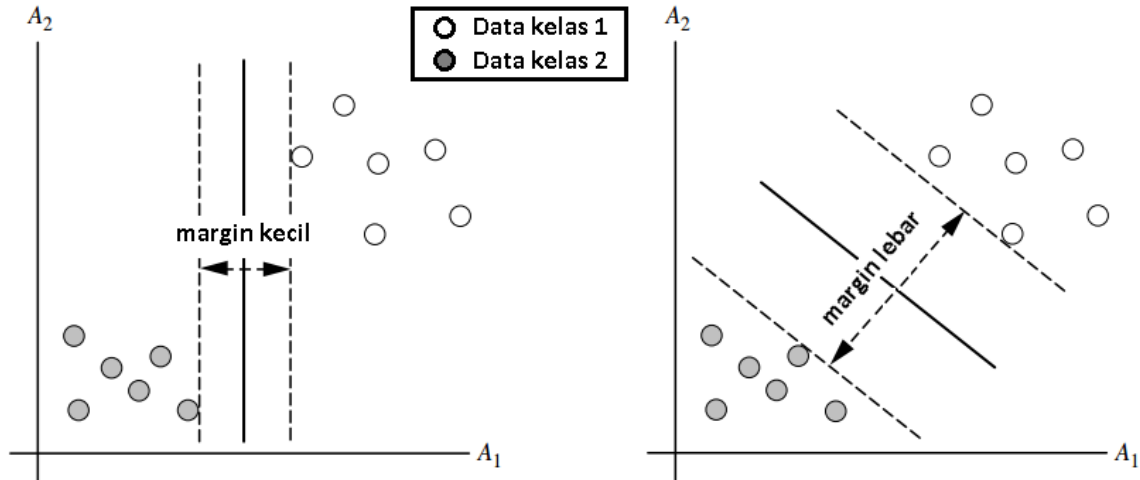
dipisahkan secara linier adalah apabila sebuah garis lurus dapat digambarkan untuk memisahkan data-data dari dua kelas yang ada. Garis lurus yang dapat memisahkan data antar dua kelas tersebut disebut dengan *hyperplane*. Namun untuk menghasilkan kesalahan klasifikasi paling kecil (minimum) maka perlu untuk menentukan satu *hyperplane* yang memisahkan data dua kelas secara optimal. Ilustrasi data dua kelas yang dapat dipisahkan oleh sebuah atau beberapa garis lurus dapat dilihat pada Gambar 2.5.



Gambar 2.5 Data dua kelas dapat dipisahkan secara linier

Algoritma SVM digunakan untuk mengatasi kesalahan klasifikasi dengan pencarian *hyperplane* dengan syarat lebar margin maksimal/terbesar (*Maximum Marginal Hyperplane* - MMH), yaitu margin yang memberikan pemisahan terbesar antar kelas. Dapat didefinisikan bahwa jarak terpendek dari *hyperplane* ke satu sisi dari *margin* adalah sama dengan jarak terpendek dari *hyperplane* ke sisi margin lainnya, dimana sisi dari margin tersebut sejajar dengan *hyperplane*. *Maximum Marginal Hyperplane* (MMH) adalah jarak terpendek dari *margin* ke data pelatihan (*training data*) terdekat dari kedua kelas.

Ilustrasi *margin* kecil dan *margin* lebar yang memisahkan data antar dua kelas seperti yang disajikan pada Gambar 2.6.



Gambar 2.6 Ilustrasi *margin* kecil dan lebar

Persamaan untuk sebuah *hyperplane* yang memisahkan dua kelas data sesuai persamaan (2.4):

$$w \cdot x + b = 0 \quad (2.4)$$

Keterangan:

$w$  : vector bobot, yang terdiri atas  $(w_1, w_2, \dots, w_m)$

$m$  : jumlah atribut

$x$  : nilai data dari masing-masing kelas yang dipisahkan yang terdiri atas  $(x_1, x_2)$ , dengan  $x_1$  adalah data kelas pertama dan  $x_2$  adalah data kelas kedua

$b$  : *scalar*/bias

Jika  $b$  dianggap sebagai bobot tambahan,  $w_0$ , maka *hyperplane* pemisah dapat dinyatakan sesuai persamaan (2.5):

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \quad (2.5)$$

Dengan demikian, setiap titik yang terletak di atas *hyperplane* yang memisahkan data kedua kelas dinyatakan sesuai persamaan (2.6):

$$w_0 + w_1 x_1 + w_2 x_2 > 0 \quad (2.6)$$

Demikian juga setiap titik yang terletak di bawah *hyperplane* yang memisahkan data kedua kelas dinyatakan sesuai persamaan (2.7):

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (2.7)$$

Bobot dapat disesuaikan agar *hyperplanes* mampu mendefinisikan sisi margin yang dinyatakan sesuai persamaan (2.8) dan (2.9):

$$H_1: w_0 + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_i = +1 \quad (2.8)$$

$$H_2: w_0 + w_1x_1 + w_2x_2 \leq -1 \text{ for } y_i = -1 \quad (2.9)$$

Keterangan:

$H_1$  : data yang berada diatas *hyperplane* yang termasuk himpunan data kelas pertama (kelas +1)

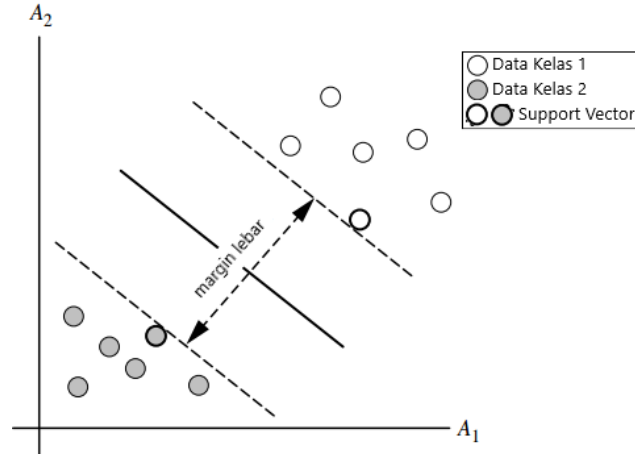
$H_2$  : data yang berada dibawah *hyperplane* yang termasuk himpunan data kelas kedua (kelas -1)

Dengan menggabungkan dua persamaan diatas, maka akan didapatkan persamaan baru yang dinyatakan seperti persamaan (2.10):

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \forall_i \quad (2.10)$$

Setiap data latih yang terletak pada *hyperplanes*  $H_1$  atau  $H_2$  (sisi margin) disebut dengan vektor pendukung (*support vector*), karena sama-sama dekat dengan *Maximum Marginal Hyperplane* (MMH) baik secara terpisah maupun berlawanan. Pada dasarnya suatu vektor pendukung adalah satu atau beberapa data yang paling sulit untuk diklasifikasikan dan mampu memberikan informasi paling banyak mengenai klasifikasi.

Vektor pendukung (*support vector*) yang terletak pada sisi margin, dapat diilustrasikan seperti pada Gambar 2.7.



Gambar 2.7 Ilustrasi *Support Vector* (vector pendukung)

Dari penjelasan di atas dapat diperoleh rumus untuk ukuran margin maksimal, yaitu jarak dari *hyperplane* ke titik manapun di  $H_1$  adalah  $\frac{1}{\|w\|}$ , dengan  $\|w\|$  adalah norma Euclidean dari  $w$ , yaitu  $\sqrt{w \cdot w}$ . Jika  $w = (w_1, w_2, \dots, w_n)$ , maka  $\sqrt{w \cdot w} = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ . Diketahui bahwa jarak dari *hyperplane* ke titik manapun di  $H_1$  sama dengan jarak dari titik mana pun pada  $H_2$  ke *hyperplane*. Maka dapat disimpulkan bahwa nilai margin maksimal adalah  $\frac{2}{\|w\|}$ . Setelah vektor dukungan dan MMH dapat diketahui, maka telah dimiliki sebuah mesin vektor dukungan (*support vector machine* – SVM) yang terlatih. MMH merupakan batas kelas linear, sehingga SVM yang diperoleh dapat digunakan untuk mengklasifikasikan data yang dapat dipisahkan secara linier. Sehingga dengan istilah lain, SVM terlatih yang dimiliki merupakan SVM linier.

Berdasarkan persamaan-persamaan yang telah dituliskan sebelumnya, MMH (*Maximum Marginal Hyperplane*) yang menjadi batas dapat dinyatakan sesuai persamaan (2.11):

$$d(x^T) = \sum_{i=1}^n y_i \alpha_i x_i x^T + b_0 \quad (2.11)$$

Keterangan:

$y_i$  : kelas/label dari *support vector*  $x_i$

$x^T$  : data tes

$\alpha_i$  dan  $b_0$  : parameter numerik yang ditentukan secara otomatis oleh optimasi atau algoritma SVM

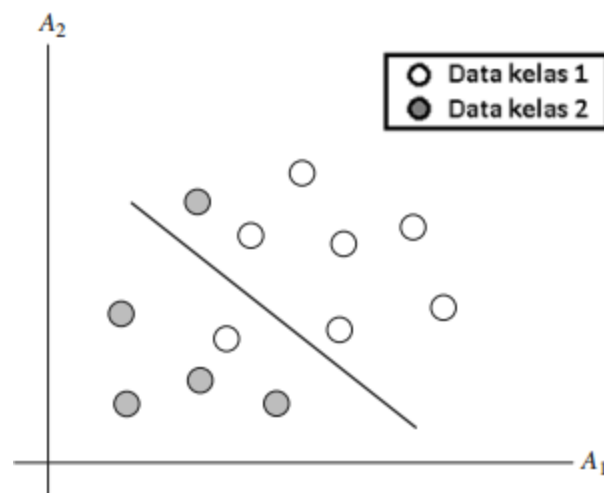
$v$  : jumlah *support vector*

Untuk data yang terpisah secara linier, vektor-vektor pendukung adalah bagian dari data pelatihan yang sebenarnya. Dengan melakukan uji data tes,  $x^T$ , pada persamaan (2.1) dapat diketahui tanda hasil yang akan memberitahu di sisi mana *hyperplane* data tes tersebut berada. Jika tanda positif, maka  $x^T$  berada di atas MMH, sehingga SVM memprediksi bahwa  $x^T$  termasuk/diklasifikasikan sebagai anggota kelas pertama (kelas +1). Jika tanda negatif, maka  $x^T$  berada di bawah MMH dan prediksi kelas adalah kelas kedua (kelas -1).

Bila diperhatikan persamaan (2.11) berisi produk titik antara vektor dukungan  $x_i$  dan data uji  $x^T$ . Hal ini membuktikan bahwa akan sangat berguna untuk mencari MMH dan vektor pendukung untuk kasus dengan data yang diberikan hanya dapat dipisahkan secara non-linier. Pada kasus non-linier, ada dua hal penting yang perlu diperhatikan. Pertama adalah kompleksitas pengklasifikasi terlatih ditandai dengan jumlah vektor pendukung yang lebih banyak daripada dimensi data. Vektor pendukung adalah data pelatihan yang penting atau kritis karena berada paling dekat dengan batas keputusan (MMH). Jika semua data pelatihan lainnya dihapus dan pelatihan diulang, maka *hyperplane* pemisah yang sama akan ditemukan. Selanjutnya, jumlah vektor pendukung yang ditemukan dapat digunakan untuk menghitung tingkat kesalahan yang diharapkan dari pengklasifikasi SVM, yang tidak bergantung pada dimensi data. Sebuah SVM dengan jumlah vektor dukungan sedikit dapat memiliki generalisasi yang baik, bahkan ketika dimensi datanya tinggi.

## B. SVM untuk data yang tidak dapat dipisahkan secara linier

Pada data yang tidak terpisah secara linier, tidak ada garis lurus yang dapat memisahkan antara dua kelas, sehingga SVM linier tidak akan dapat menemukan solusi yang layak. Namun pendekatan pada SVM linier dapat diperluas guna membuat SVM nonlinier untuk klasifikasi data yang tidak dapat dipisahkan secara linier, juga disebut data terpisah nonlinier atau data nonlinier. SVM tersebut mampu menemukan batas keputusan nonlinier, yaitu *hypersurfaces* nonlinier di ruang input. Ilustrasi data dua kelas yang tidak dapat dipisahkan dengan sebuah garis lurus seperti pada Gambar 2.8.



Gambar 2.8 Data dua kelas tidak dapat dipisahkan secara linier

SVM nonlinier dapat diperoleh dengan memperluas pendekatan untuk SVM linier dengan dua langkah utama. Pada langkah pertama, dengan mengubah data input asli ke dalam ruang dimensi yang lebih tinggi menggunakan pemetaan nonlinier. Setelah data diubah menjadi ruang baru yang lebih tinggi, langkah kedua adalah mencari *hyperplane* pemisah linear di ruang baru. Lalu dengan menggunakan formulasi SVM linier untuk menyelesaikan masalah optimasi kuadrat yang ditemukan. MMH yang ditemukan di ruang baru akan menyesuaikan dengan *hypersurface* pemisahan nonlinear di ruang asli.



Transformasi nonlinier dari data input asli ke ruang dimensi yang lebih tinggi, dapat dijelaskan dengan contoh berikut. Pada kasus linier, sebuah *hyperplane* keputusan di ruang baru adalah  $d(z) = w \cdot z + b$ , dengan  $w$  dan  $z$  adalah vektor pendukung. Misalkan vektor  $X = (x_1, x_2, x_3)$  pada ruang 3D, dipetakan sebagai vektor  $Z$  pada ruang 6D, dengan menggunakan pemetaan  $\phi_1(X) = x_1$ ,  $\phi_2(X) = x_2$ ,  $\phi_3(X) = x_3$ ,  $\phi_4(X) = (x_1)^2$ ,  $\phi_5(X) = x_1x_2$ , dan  $\phi_6(X) = x_1x_3$ . Perlu untuk memecahkan  $w$  dan  $b$  dan kemudian mengganti kembali sehingga *hyperplane* keputusan linier dalam ruang ( $Z$ ) baru sesuai dengan polinom orde kedua nonlinier di ruang input 3-D asli, dapat dinyatakan sesuai persamaan (2.12):

$$\begin{aligned} d(Z) &= w_1x_1 + w_2x_2 + w_3x_3 + w_4(x_1)^2 + w_5x_1x_2 + w_6x_1x_3 + b \quad (2.12) \\ &= w_1z_1 + w_2z_2 + w_3z_3 + w_4z_4 + w_5z_5 + w_6z_6 + b \end{aligned}$$

Namun terdapat beberapa masalah, diantaranya adalah bagaimana kita memilih pemetaan nonlinier ke ruang dimensi yang lebih tinggi, dan perhitungan yang terlibat akan menjadi mahal. Untuk mengatasi permasalahan tersebut, maka digunakan trik matematika. Sebagai alternatif daripada menghitung produk titik pada data yang diubah, ternyata secara matematis setara dengan penerapan fungsi kernel  $K$  yaitu  $K = (x_i, x_j)$ , ke input data asli, sehingga dapat dinyatakan sesuai persamaan (2.13):

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (2.13)$$

Diibaratkan bahwa setiap  $\phi(x_i) \cdot \phi(x_j)$  muncul dalam persamaan, maka akan dapat diganti dengan  $K(x_i, x_j)$ . Dengan cara ini, semua perhitungan dibuat di ruang input asli, sehingga berpotensi memiliki dimensi yang lebih rendah yang mampu menghindari pemetaan dengan aman. Setelah penerapan trik matematika ini, maka dapat dilanjutkan untuk pencarian *hyperplane* pemisah maksimal, dengan melibatkan batas atas yang ditentukan secara experimentally.

Terdapat tiga fungsi kernel yang dapat diterima untuk digunakan guna menggantikan skenario produk titik yang telah dijelaskan, diantaranya adalah kernel polinomial derajat  $h$  sesuai persamaan (2.14), kernel *Gaussian Radial Basis Function*-RBF sesuai persamaan (2.15), dan kernel Sigmoid sesuai persamaan (2.16):

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^h \quad (2.14)$$

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \quad (2.15)$$

$$K(x_i, x_j) = \tanh(\kappa x_i \cdot x_j - \delta) \quad (2.16)$$

Masing-masing kernel tersebut merupakan hasil dalam *classifier* nonlinier yang berbeda dalam ruang input. Dapat dicatat bahwa keputusan yang dihasilkan dari *hyperplanes* yang ditemukan untuk SVM nonlinier adalah tipe yang sama seperti yang ditemukan oleh pengklasifikasi jaringan saraf lainnya. Misalnya, SVM dengan *Gaussian Radial Bases Function* (RBF) memberikan *hyperplane* keputusan yang sama sebagai jenis jaringan saraf yang dikenal sebagai jaringan fungsi basis radial (*Radial Bases Function* - RBF). SVM dengan *kernel sigmoid* setara dengan jaringan neural dua-lapisan sederhana yang dikenal sebagai *multilayer perceptron*. Tidak ada aturan terbaik untuk menentukan kernel yang diterima akan menghasilkan SVM yang paling akurat. Dalam prakteknya, kernel yang dipilih umumnya tidak membuat perbedaan besar dalam menghasilkan akurasi. Pelatihan SVM selalu menemukan solusi global.

Pengklasifikasi SVM dapat dikombinasikan untuk kasus multi kelas (*multiclass*). Pendekatan yang sederhana dan efektif, diberikan kepada multi kelas sebanyak  $m$ , melatih pengklasifikasi sebanyak  $m$ , untuk setiap kelas, di mana *classifier*  $j$  belajar untuk mengembalikan nilai positif untuk kelas  $j$  dan nilai negatif untuk sisanya. Data tes yang diberikan kelas disesuaikan dengan jarak positif terbesar.

SVM juga dapat dirancang untuk regresi linier dan nonlinier. SVM untuk regresi mampu mempelajari hubungan input-output antara data pelatihan input, dan output bernilai kontinyu yang sesuai. Tujuan penelitian utama mengenai SVM adalah untuk meningkatkan kecepatan dalam pelatihan dan pengujian sehingga SVM dapat menjadi pilihan yang lebih layak untuk kumpulan data yang sangat besar (misalnya jutaan vektor dukungan), juga masalah lain termasuk menentukan kernel terbaik untuk kumpulan data tertentu dan menemukan metode yang lebih efisien untuk kasus *multiclass*.

### 3. NB (*Naive Bayes*)

*Naive Bayes* (Cherman et al., 2011) menggunakan teorema Bayes tetapi tidak memperhitungkan saling ketergantungan antar akun yang mungkin ada antara fitur/atribut. Klasifikasi *Naive Bayes* (Sayad, 2010-2018c) didasarkan pada teorema Bayes dengan asumsi independensi antara prediktor/atribut/fitur. Model *Naive Bayes* mudah dibuat, tanpa estimasi parameter iteratif yang rumit dan sangat berguna untuk dataset besar. Meskipun sederhana, pengklasifikasi *Naive Bayes* secara mengejutkan dalam banyak penggunaan sering memberikan hasil lebih baik dibanding metode klasifikasi lain yang lebih canggih.

Pengklasifikasi Bayesian adalah penggolongan statistik, dan pengklasifikasi Bayesian yang sederhana dikenal juga dengan pengklasifikasi *Naive Bayes* (Han et al., 2012). Pengklasifikasi *Naive Bayes* yang didasarkan pada teorema Bayes dapat memprediksi probabilitas keanggotaan kelas seperti probabilitas bahwa data uji yang diberikan adalah milik kelas tertentu, dan juga telah menunjukkan akurasi dan kecepatan yang tinggi ketika diterapkan ke database yang sangat besar. Pengklasifikasi *Naive Bayes* menganggap bahwa efek dari nilai atribut pada kelas tertentu tidak bergantung pada nilai atribut lainnya. Asumsi ini disebut independensi kelas bersyarat. Pengklasifikasi *Naive Bayes* disebut “naif (*naive*)” karena dibuat untuk menyederhanakan perhitungan yang digunakan. Nama teorema Bayes diambil dari Thomas

Bayes, seorang pendeta Inggris yang memulai penelitian awal dalam teori probabilitas dan keputusan selama abad ke-18.

Dalam istilah Bayesian,  $A$  adalah kumpulan data yang dianggap sebagai "bukti" yang merupakan satu set atribut sebanyak  $m$ . Lalu  $H$  adalah hipotesis bahwa data  $A$  termasuk kelompok kelas  $C$ . Untuk melakukan klasifikasi, lebih dahulu dengan menentukan  $P(H|A)$ , yaitu probabilitas bahwa hipotesis  $H$  memegang "bukti" atau data yang diamati  $A$ . Dengan kata lain adalah mencari kemungkinan bahwa data  $A$  milik kelas  $C$ , dengan deskripsi atribut  $A$  telah diketahui.  $P(H|A)$  adalah probabilitas posterior/posteriori dari hipotesis  $H$  atas kondisi data  $A$ . Sebaliknya  $P(H)$  adalah probabilitas prior/priori dari  $H$ . Probabilitas posterior  $P(H|A)$  didasarkan pada informasi daripada probabilitas prior  $P(H)$ , namun tidak bergantung (*independent*) pada  $A$ . Dapat juga disamakan untuk  $P(A|H)$ , probabilitas posterior data  $A$  atas kondisi hipotesis  $H$ . Diketahui  $P(A)$  adalah probabilitas prior dari  $A$ . Probabilitas  $P(H)$ ,  $P(A|H)$ , dan  $P(A)$  dapat dihitung dari data yang diperoleh. Teorema Bayes memberikan cara untuk menghitung probabilitas  $P(H|A)$ , dari  $P(H)$ ,  $P(A|H)$ , dan  $P(A)$ , yang dinyatakan sesuai persamaan (2.17):

$$P(H|A) = \frac{P(A|H)P(H)}{P(A)} \quad (2.17)$$

Langkah penggunaan pengklasifikasi *Naive Bayes* yang merupakan pengklasifikasi Bayesian sederhana, sebagai berikut:

- 1) Bila  $D$  adalah satu set data pelatihan dengan kelas label terkaitnya, setiap data diwakili oleh atribut dengan  $m$  dimensi,  $A = (x_1, x_2, \dots, x_m)$ , dengan  $m$  pengukuran pada data dari  $m$  digambarkan sebagai  $A_1, A_2, \dots, A_m$ .
- 2) Kelas sejumlah  $m$  digambarkan sebagai  $C_1, C_2, \dots, C_m$ . Bila diberikan data tes  $A$ , pengklasifikasi *Naive Bayes* akan memberikan hasil prediksi  $A$  adalah termasuk pada kelas yang memiliki probabilitas posterior tertinggi, berdasarkan kondisi  $A$ . Pengklasifikasi *Naive Bayes*

memprediksi bahwa data uji  $X$  milik kelas  $C_i$  jika dan hanya jika sesuai persamaan (2.18):

$$P(C_i/A) > P(C_j/A) \quad \text{untuk } 1 \leq j \leq m, \quad j \neq i \quad (2.18)$$

Kelas  $C_i$  dengan memaksimalkan  $P(C_i/A)$  disebut hipotesis posteriori maksimum (*maximum posteriori hypothesis*), yang dengan teorema Bayes dapat dinyatakan sesuai persamaan (2.19):

$$P(C_i/A) = \frac{P(A/C_i)P(C_i)}{P(A)} \quad (2.19)$$

Keterangan:

$P(C_i/A)$  : probabilitas posterior kelas (target) atas prediktor (atribut) yang diberikan

$P(A/C_i)$  : kemungkinan yang merupakan probabilitas dari prediktor yang diberikan kelas

$P(C_i)$  : probabilitas prior dari kelas

$P(A)$  : probabilitas prior dari prediktor

- 3) Jika  $P(A)$  adalah konstan untuk semua kelas, maka  $P(A/C_i)P(C_i)$  perlu dimaksimalkan. Jika probabilitas prior kelas tidak diketahui, maka diasumsikan semua kelas memiliki kemungkinan/probabilitas yang sama,  $P(C_1) = P(C_2) = \dots = P(C_m)$ , sehingga  $P(A/C_i)$  yang perlu untuk dimaksimalkan. Namun untuk memaksimalkan  $P(A/C_i)P(C_i)$  perlu diperhatikan bahwa probabilitas prior kelas dapat diperkirakan sebagai  $P(C_i) = |C_{i,D}|/|D|$ , dengan  $|C_{i,D}|$  adalah jumlah data pelatihan pada kelas  $C_i$  dalam kumpulan data  $D$ .
- 4) Untuk kumpulan data yang memiliki banyak atribut, akan sangat mahal untuk melakukan komputasi guna menghitung  $P(A/C_i)$ . Untuk mengurangi kompleksitas perhitungan  $P(A/C_i)$ , maka dibuatlah asumsi naif untuk kebebasan kelas bersyarat, yaitu dengan mengibaratkan bahwa tidak ada ketergantungan antar atribut sesuai persamaan (2.20):

$$\begin{aligned} P(A/C_i) &= \prod_{k=1}^n P(x_k/C_i) \\ &= P(x_1/C_i) * P(x_2/C_i) * \dots * P(x_n/C_i) \end{aligned} \quad (2.20)$$

Sehingga dapat diperkirakan dengan mudah probabilitas  $P(x_1|C_i) * P(x_2|C_i) * ... * P(x_n|C_i)$  dari data pelatihan. Dengan  $x_k$  adalah nilai atribut  $A_k$  dari data  $A$ . Setiap atribut akan di cek apakah bersifat kategorikal atau bernilai kontinyu. Untuk menghitung akan dipertimbangkan beberapa hal, sebagai berikut:

- a. Jika  $A_k$  bersifat kategorikal, maka  $P(x_k|C_i)$  adalah jumlah data pada kelas  $C_i$  dalam kumpulan data  $D$  yang mempunyai nilai  $x_k$  untuk  $A_k$ , dibagi dengan  $|C_{i,D}|$  (jumlah data pada kelas  $C_i$  dalam kumpulan data  $D$ ).
- b. Jika  $A_k$  bernilai kontinyu, maka atribut tersebut diasumsikan berdistribusi *Gaussian* yang memiliki nilai rerata ( $\mu$ ) dan standar deviasi ( $\sigma$ ) sesuai persamaan (2.21):

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.21)$$

Sehingga dapat disederhanakan seperti yang dinyatakan pada persamaan (2.22):

$$P(x_k|C_i) = g(x_k, \mu_i, \sigma_{C_i}) \quad (2.22)$$

- 5) Untuk memprediksi kelas label  $X$ , maka nilai  $P(A|C_i)P(C_i)$  dihitung untuk setiap kelas  $C_i$ . Pengklasifikasi akan memberikan hasil prediksi kelas label dari data  $A$  adalah kelas  $C_i$  dengan  $P(A|C_i)P(C_i)$  yang terbesar/maksimal, adalah jika dan hanya jika sesuai persamaan (2.23):

$$P(A|C_i)P(C_i) > P(A|C_j)P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (2.23)$$

Secara teori, pengklasifikasi *Naive Bayes* memiliki tingkat kesalahan minimum dibandingkan dengan pengklasifikasi lainnya. Namun, dalam prakteknya ini tidak selalu terjadi, karena ketidakakuratan dalam asumsi yang dibuat untuk penggunaannya, seperti independensi syarat kelas, dan kurangnya data probabilitas yang tersedia. Sebagai contoh, di bawah asumsi tertentu, pengklasifikasi ini dapat menunjukkan

bahwa banyak jaringan saraf dan algoritma kurva-pas menghasilkan hipotesis posteriori maksimum, seperti pengklasifikasi *Naive Bayes*.

#### 4. *Random Forest* (RF)

*Random Forest* (RF) (Cutler et al., 2012) diperkenalkan oleh Leo Breiman, merupakan algoritma yang dapat digunakan untuk variabel respon kategorikal dalam melakukan klasifikasi atau regresi, dan variabel prediktor dapat berupa tipe kategori atau nilai kontinyu. Hal-hal yang menjadikan RF menarik dari segi komputasi/perhitungan adalah sebagai berikut:

- secara alami dapat menangani klasifikasi, regresi dan klasifikasi *multiclass*;
- relatif cepat untuk dilatih dan diprediksi;
- hanya bergantung pada satu atau dua parameter penyetelan;
- memiliki estimasi kesalahan generalisasi yang dibangun;
- dapat digunakan secara langsung untuk masalah-masalah berdimensi tinggi;
- dapat dengan mudah diimplementasikan secara paralel.

Sedangkan hal-hal yang menjadikan RF menarik secara statistik adalah karena fitur-fitur tambahan yang tertanam, sebagai berikut:

- ukuran-ukuran kepentingan variabel;
- pembobotan kelas diferensial;
- tidak ada nilai yang hilang;
- visualisasi;
- deteksi pencilan;
- pembelajaran tanpa pengawasan.

Sesuai namanya, RF adalah pengembangan berbasis pohon, dengan setiap pohon bergantung pada kumpulan variabel acak. Untuk lebih formalnya, yaitu vektor acak dengan dimensi  $p$  yaitu  $A = (A_1, A_2, \dots, A_p)^T$  yang mewakili nilai nyata masukan atau variabel prediktor, dan sebuah

variabel acak  $L$  mewakili nilai nyata respon/tanggapan, sehingga dapat diperkirakan distribusi bersama sebagai  $P_{A,L}(A, L)$ . Tujuannya adalah untuk menemukan fungsi prediksi  $f(A)$  guna memprediksi  $L$ . Fungsi prediksi tersebut ditentukan oleh fungsi kerugian (*loss*)  $Loss(L, f(A))$ , dan untuk meminimalkan nilai kerugian yang diperkirakan didefinisikan sesuai persamaan (2.24):

$$E_{A,L}(Loss(L, f(A))) \quad (2.24)$$

Secara intuitif,  $Loss(L, f(A))$  mengukur seberapa dekat  $f(A)$  dengan  $L$ , dengan mengabaikan nilai  $f(A)$  yang sangat jauh dengan  $L$ . Pilihan umum dari  $Loss$  (fungsi kerugian) adalah *squared error loss*  $Loss(L, f(A)) = (L - f(A))^2$  untuk regresi dan *zero-one loss* untuk klasifikasi yang dapat dinyatakan sesuai persamaan (2.25):

$$Loss(L, f(A)) = I(L \neq f(A)) = \begin{cases} 0 & \text{jika } L = f(A) \\ 1 & \text{untuk lainnya} \end{cases} \quad (2.25)$$

Namun untuk meminimalkan  $E_{A,L}(Loss(L, f(A)))$  pada *squared error loss*, dibutuhkan syarat ekspektasi yang juga dikenal sebagai fungsi regresi sesuai persamaan (2.26):

$$f(x) = E(L/A=x) \quad (2.26)$$

Dalam masalah klasifikasi, jika himpunan nilai yang mungkin dari  $L$  dilambangkan oleh  $\mathcal{Y}$ , maka untuk meminimalkan  $E_{A,L}(Loss(L, f(A)))$  pada *zero-one loss* berlaku syarat yang dikenal sebagai aturan Bayes yang didefinisikan sesuai persamaan (2.27):

$$f(x) = \operatorname{argmax}_y P(L=y/A=x), y \in \mathcal{Y} \quad (2.27)$$

Konstruksi pengembangan  $f$  dalam bentuk kumpulan disebut “*base learner*”  $h_1(x), h_2(x), \dots, h_J(x)$ , dan dapat dikombinasikan dengan pengembangan prediktor  $f(x)$ . Pada masalah regresi *base learner* dirata-ratakan sehingga dapat didefinisikan sesuai persamaan (2.28):

$$f(x) = \frac{1}{J} \sum_{j=1}^J h_j(x) \quad (2.28)$$



Sedangkan pada masalah klasifikasi,  $f(x)$  adalah kelas yang paling sering diprediksi (*voting*) sesuai yang didefinisikan pada persamaan (2.29):

$$f(x) = \operatorname{argmax}_I \sum_{j=1}^J I(y=h_j(x)), y \in \mathcal{Y} \quad (2.29)$$

Dalam RF, *base learner* ke- $j$  adalah pohon yang dinotasikan dengan  $h_j(X, \Theta_j)$ , di mana  $\Theta_j$  adalah kumpulan variabel acak dan independen untuk  $j = 1, 2, \dots, J$ . Meskipun definisi dari RF sangat umum, namun hampir selalu diimplementasikan dengan cara tertentu. Untuk memahami algoritma RF, penting untuk memiliki pengetahuan dasar tentang jenis pohon (*trees*) yang digunakan sebagai *base learner*.

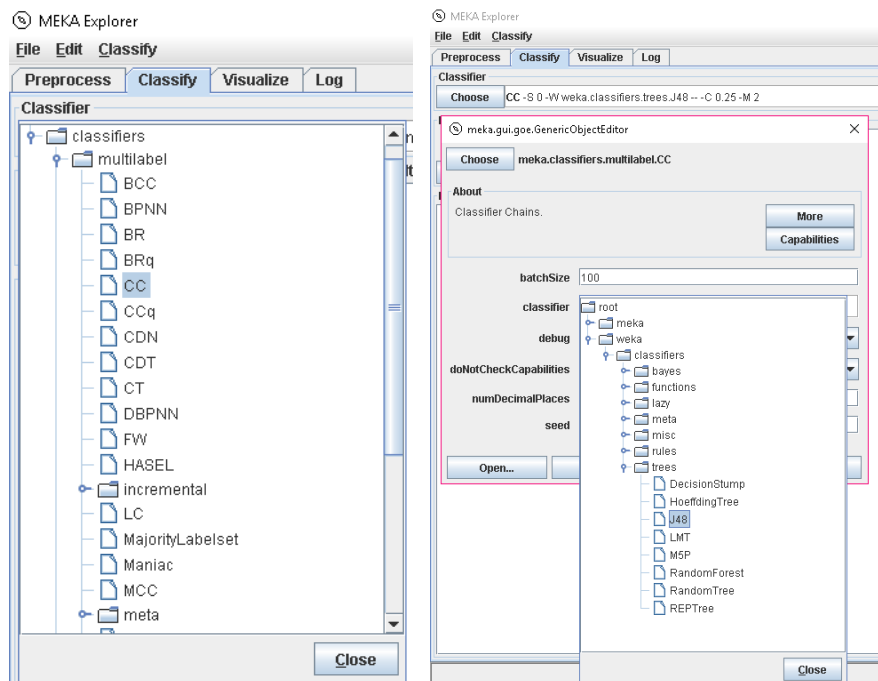
RF dapat dikatakan sebagai alat multiguna, karena dapat digunakan untuk masalah regresi dan klasifikasi, termasuk klasifikasi multi kelas. RF memberikan perkiraan internal kesalahan generalisasi sehingga validasi silang tidak diperlukan. RF dapat disetel, namun sering bekerja cukup baik dengan parameter pengaturan *default*. Variabel ukuran penting yang dapat digunakan untuk pemilihan variabel juga tersedia. RF menghasilkan proximities (kedekatan) untuk menghitung nilai yang hilang. Proximitas juga dapat memberikan banyak informasi dengan memungkinkan visualisasi baru dari data. RF telah berhasil digunakan untuk berbagai macam aplikasi sehingga menjadi populer untuk digunakan dalam beberapa disiplin ilmu.

#### 2.4.5 MEKA 1.9.1

Salah satu alat (*tool*) untuk melakukan Klasifikasi Multi Label yaitu aplikasi MEKA yang merupakan pengembangan dari WEKA. WEKA adalah alat (*tool*) yang dikembangkan untuk melakukan klasifikasi label tunggal, sedangkan MEKA merupakan perpanjangan multi-label/multitarget dari WEKA (Read et al., 2016). MEKA dibuat untuk melakukan dan mengevaluasi Klasifikasi Multi Label menggunakan Metode Transformasi Masalah yang populer dan efektif, yang juga menggunakan metode klasifikasi label tunggal sebagai pengklasifikasi dasar. MEKA yang digunakan pada penelitian ini adalah versi 1.9.1 yang dirilis pada 12

April 2017 (Read, 2012-2017), dengan beberapa perbaikan dari versi sebelumnya dan juga mencakup fitur baru dan pengklasifikasi baru.

Terdapat banyak pengklasifikasi (*classifier*) yang didukung oleh MEKA v.1.9.1 untuk Metode Transformasi Masalah, yang merupakan salah satu pendekatan Klasifikasi Multi Label. Pada penggunaan MEKA, pengklasifikasi Multi Label yang ada dikombinasikan dengan pengklasifikasi dasar (*base classifier*). Pengklasifikasi dasar tersebut sama dengan yang didukung oleh WEKA. Daftar pengklasifikasi Multi Label dan pengklasifikasi dasar yang didukung oleh MEKA v.1.9.1 seperti yang ditampilkan pada Gambar 2.9.

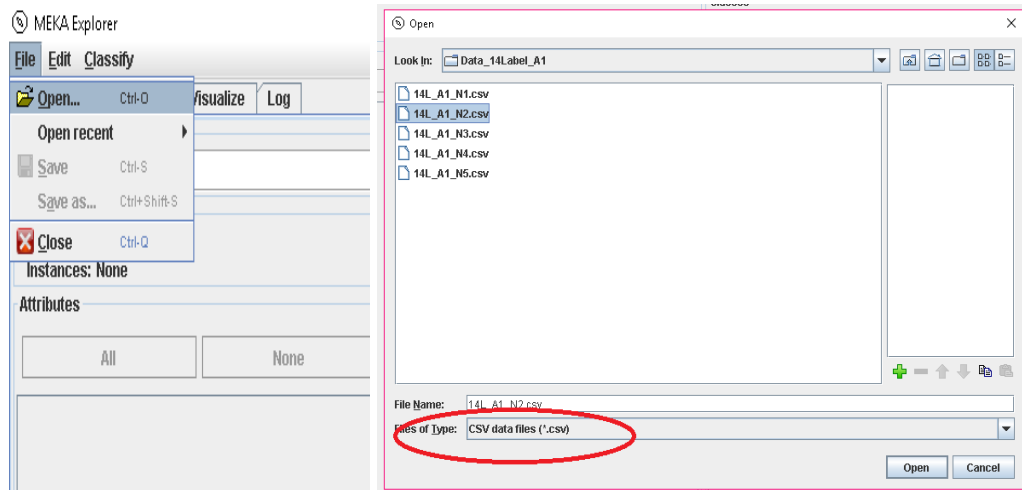


Gambar 2.9 Pengklasifikasi Multi Label dan dasar pada MEKA v.1.9.1

#### 2.4.6 Penggunaan MEKA

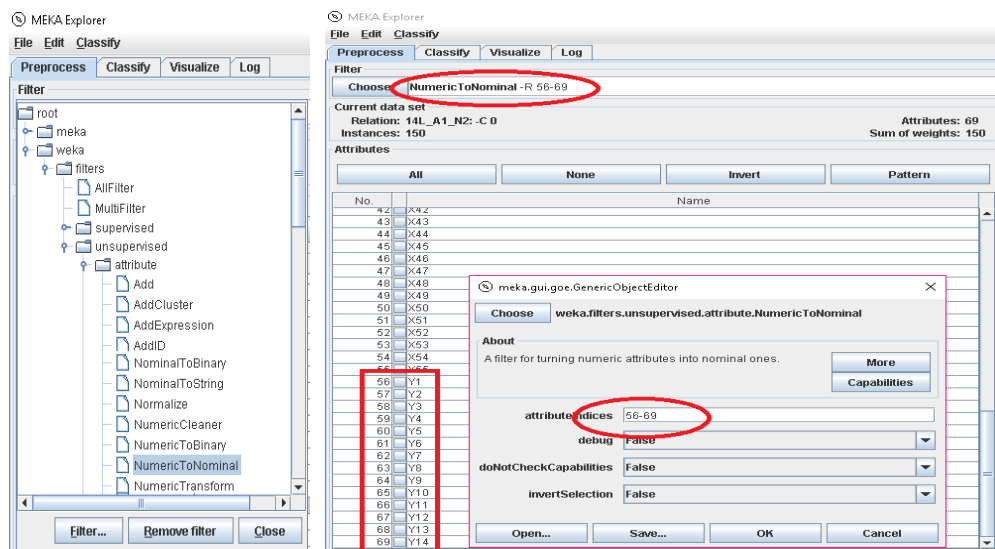
Langkah-langkah dalam melakukan pengujian terhadap dataset (*instance*) dengan menggunakan Metode Transformasi Masalah (*Problem Transformation Methods*), yang merupakan salah satu pendekatan Klasifikasi Multi Label (*multi label classification*), pada aplikasi MEKA versi 1.9.1 yaitu menggunakan kombinasi pengklasifikasi multi label dengan pengklasifikasi dasar adalah sebagai berikut:

1. Membuka file dataset dalam format csv yang akan diujikan dengan MEKA seperti yang ditampilkan pada Gambar 2.10;



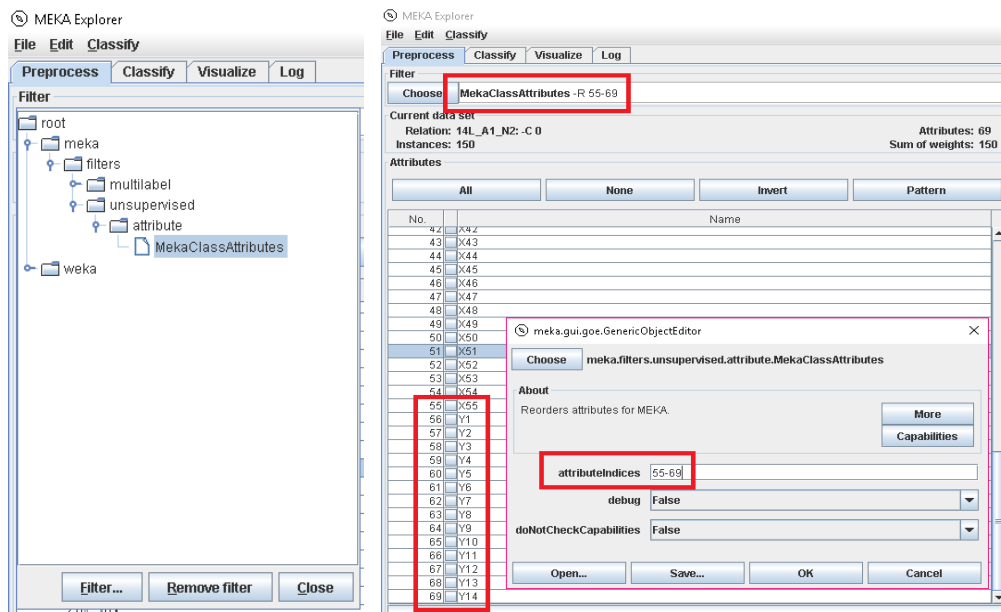
Gambar 2.10 Buka File Dataset dalam format CSV

2. Merubah tipe data dari *numeric* menjadi *nominal* untuk semua data Label, yaitu dengan memilih *filter numerictonominal* pada *tab Preprocess* dengan urutan pilihan pada Filter adalah “*root-weka-filters-unsupervised-attribute-NumericToNominal*” dan menentukan data label yang akan dirubah, sedangkan untuk data atribut dibiarkan tetap dengan tipe data *numeric*, seperti yang ditampilkan pada Gambar 2.11;



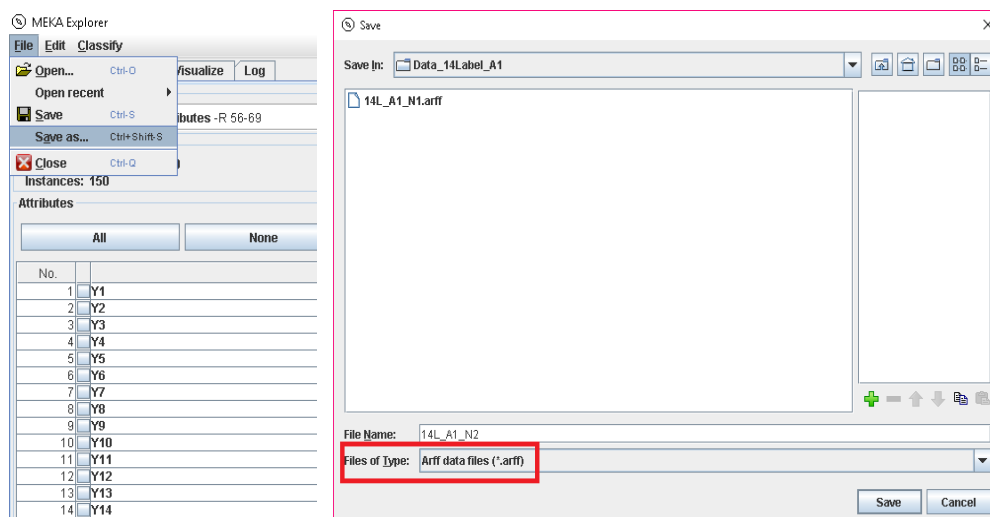
Gambar 2.11 Rubah Tipe Data *Numeric* ke *Nominal* untuk data Label

- Menentukan data yang akan dijadikan Label dengan memilih filter *MekaClassAttributes* dengan urutan pilihan pada Filter adalah “*root-meka-filters-unsupervised-attribute-MekaClassAttribute*”, seperti yang ditampilkan pada Gambar 2.12;



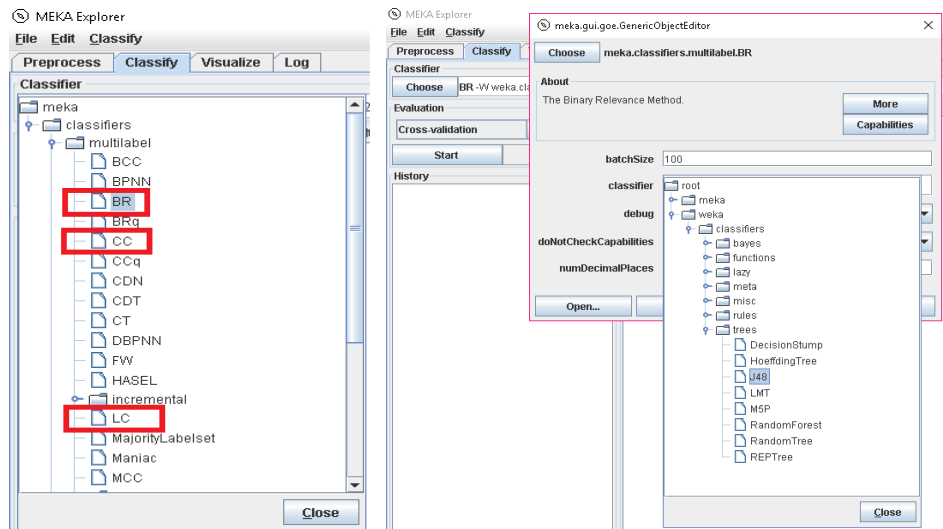
Gambar 2.12 Menentukan Label untuk Klasifikasi

- Menyimpan dataset dalam format arff yang didukung oleh MEKA, seperti yang ditampilkan pada Gambar 2.13;



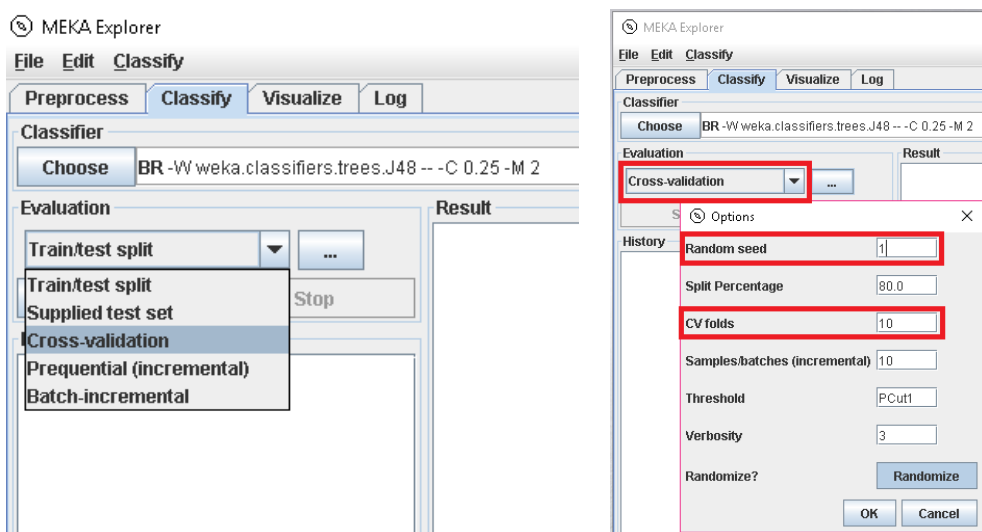
Gambar 2.13 Simpan dataset dalam format ARFF

5. Memilih *classifier* yang akan diujikan pada dataset, yaitu dengan memilih *classifier* multi label dan kombinasi *base classifier* yang akan digunakan, seperti yang ditampilkan pada Gambar 2.14;



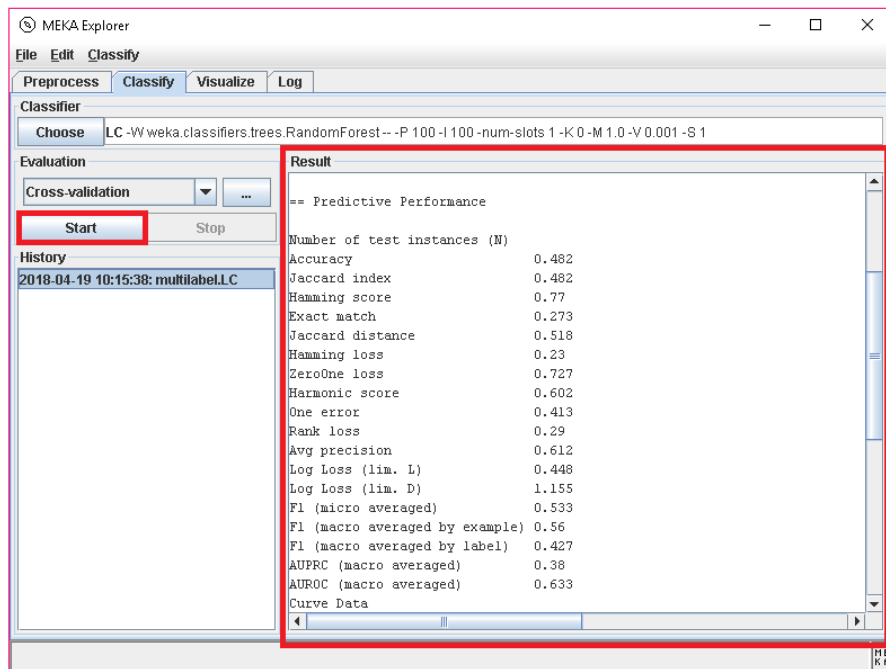
Gambar 2.14 Memilih Classifier yang akan digunakan

6. Memilih metode evaluasi *Cross-validation* pada *tab Classify*, dengan merubah setingan pada *Random Seed* (1 s.d 10), dan *CVfolds* (10) sedangkan setingan/pengaturan lain dibiarkan sesuai awal (*default*), seperti yang ditampilkan pada Gambar 2.15;



Gambar 2.15 Pilih Metode Evaluasi

- Setelah melakukan klasifikasi dengan menekan tombol “Start”, maka MEKA akan menampilkan nilai pengukuran hasil pengujian atas pilihan *classifier-base classifier* yang diujikan pada dataset yang dimasukan (*input*), seperti yang ditampilkan pada Gambar 2.16.



Gambar 2.16 Hasil ditampilkan oleh MEKA

#### 2.4.7 Matrik Evaluasi (Evaluation Metrics)

Terdapat banyak pengukuran untuk menghitung hasil performa/kinerja Klasifikasi Multi Label (Herrera et al., 2016; Read et al., 2015; Tsoumakas et al., 2010), diantaranya adalah sebagai berikut:

- Hamming Loss* (HL): mengevaluasi berapa kali label *instance* salah dikategorikan, yaitu label yang bukan milik *instance* diprediksi atau label milik *instance* tidak diprediksi. Semakin kecil nilai HL semakin baik kinerjanya, kinerja maksimal/sempurna adalah bila nilai HL adalah 0 (nol). HL merupakan metrik evaluasi pengukuran kinerja/performa yang paling umum digunakan dalam klasifikasi multi label (*multi label classification-MLC*). HL menghitung perbedaan simetris antara  $L_i$  (label aktual ke- $i$ ), dan  $\bar{L}_i$  (label hasil prediksi ke- $i$ ). Dengan kata lain HL menghitung jumlah

label yang salah diprediksi. Jumlah total kesalahan dalam *instance* (data) dikumpulkan dan kemudian dinormalisasi dengan mempertimbangkan jumlah label dan jumlah *instance*. Karena penghitungan kesalahan prediksi label dibagi dengan jumlah label, maka metrik ini akan menghasilkan penilaian yang berbeda untuk jumlah kesalahan yang sama ketika digunakan dengan dataset multi label (*multi label dataset*-MLD) yang memiliki jumlah label berbeda. Oleh karena itu, metrik ini merupakan indikator kesalahan yang sebanding dengan jumlah label. Perhitungan *Hamming Loss* dinyatakan sesuai persamaan (2.30):

$$HL = \frac{1}{n} \frac{1}{k} \sum_{i=1}^n |L_i \Delta \bar{L}_i| \quad (2.30)$$

Keterangan:

- $k$  : jumlah label
- $n$  : jumlah data (*instance*)
- $L_i$  : label aktual data ke- $i$
- $\bar{L}_i$  : label hasil prediksi data ke- $i$

2. *One-Error* (OE): menghitung berapa kali label peringkat teratas tidak ada di kumpulan label yang tepat dari *instance*. Nilai sempurna *One-Error* adalah 0 (nol), semakin kecil nilainya menandakan kinerja yang lebih baik. *One-Error* adalah metrik pengukuran kinerja untuk meminimalkan. Ekspresi yang mengikuti penjumlahan menghasilkan 1 jika label peringkat teratas dalam prediksi yang diberikan oleh penggolong tidak termasuk dalam label aktual. Jumlah kesalahan prediksi terakumulasi (dijumlahkan) dan dirata-ratakan. Hasilnya adalah persentase kasus di mana label yang dihasilkan pengklasifikasi (*classifier*) adalah *False Positif* (FP). Pada masalah klasifikasi label tunggal, *One-Error* identik dengan kesalahan klasifikasi biasa (*ordinary classification error*). Perhitungan *One-Error* dinyatakan sesuai persamaan (2.31):

$$OE = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\text{argmax}\langle \text{rank}(x_i, y) \rangle \notin L_i], y \in \bar{L}_i \quad (2.31)$$

Keterangan:

$x_i$  : data ke i

$argmax$  : *arguments of the maxima*, fungsi yang menentukan nilai input untuk mendapatkan hasil nilai output sebesar mungkin (maksimal)

3. *Rank Loss* (RL): mengevaluasi berapa kali label prediksi yang tidak relevan memiliki peringkat lebih tinggi daripada label yang relevan. Kinerjanya sangat bagus pada nilai 0 (nol), kinerja lebih baik jika nilainya semakin kecil.

RL mengambil semua kemungkinan kombinasi label yang relevan dan tidak relevan untuk sebuah *instance* dan menghitung berapa kali label yang tidak relevan berada di peringkat di atas label yang relevan dalam hasil prediksi. Penghitungannya dinormalisasi dengan membagi hasil prediksi dari label yang relevan dan tidak relevan dalam *instance* dan kemudian dirata-ratakan dengan jumlah *instance* yang dinilai. Perhitungan *Rank Loss* dinyatakan sesuai persamaan (2.32):

$$RL = \frac{1}{n} \sum_{i=1}^n \frac{1}{|L_i| \cdot |\bar{L}_i|} |y_a, y_b : rank(x_i, y_a) > rank(x_i, y_b), (y_a, y_b) \in L_i \times \bar{L}_i| \quad (2.32)$$

4. *Average Precision* (AP): mengevaluasi fraksi rata-rata label yang relevan memiliki peringkat yang lebih tinggi dari satu label relevan lainnya. Kinerjanya sangat bagus pada nilai 1 (nol), semakin besar nilainya semakin baik kinerjanya.

Dalam persamaan berikut, peringkat ( $x_i$ , l) didefinisikan sebagai fungsi yang untuk instance  $x_i$  dan label yang relevan  $\in Y$ , yang posisinya diketahui, mengembalikan derajat kepercayaan saya ke dalam prediksi  $Z_i$  yang dikembalikan oleh penggolong. *Average precision* menentukan proporsi label yang relevan yang diberi peringkat di atasnya dalam peringkat yang diprediksi untuk setiap label dalam suatu *instance*. Tujuannya adalah untuk mengetahui berapa rata-rata banyak posisi yang



harus diperiksa, sebelum label yang tidak relevan ditemukan. Perhitungan *Average Precision* dinyatakan sesuai persamaan (2.33):

$$AP = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\{y' | \text{rank}(x_i, y') \leq \text{rank}(x_i, y), y' \in Y_i\}|}{\text{rank}(x_i, y)} \quad (2.33)$$

## 2.5 Normalisasi Data

Praproses (*preprocessing*) data meliputi persiapan data, diantaranya adalah integrasi, pembersihan, normalisasi, dan transformasi data, dan pengurangan data seperti pemilihan fitur, pemilihan contoh, dan diskritisasi. Hasil yang diharapkan setelah rangkaian *preprocessing* data adalah dataset akhir, yang dapat dianggap benar dan berguna untuk algoritma penambangan data lebih lanjut. Persiapan data meliputi beberapa tugas yang beragam, termasuk pembersihan dan normalisasi data, berkaitan dengan nilai-nilai yang hilang, mengatasi *noise*, dan deteksi nilai ekstrim. (García et al., 2015)

Normalisasi data yang paling umum diterapkan pada atribut dengan tipe data numerik. Tujuannya adalah untuk memudahkan proses pembelajaran algoritma penambangan data, dan menormalkan skala nilai. Normalisasi adalah teknik transformasi sederhana yang tujuannya adalah untuk menyesuaikan nilai-nilai atribut untuk berbagi skala umum, membuat atribut menjadi proporsional dan sebanding. (Herrera et al., 2016)

Normalisasi pada penambangan data (*Data Mining*) adalah sebuah proses untuk penskalaan sehingga nilai atribut dari data berada pada *range* (rentang) tertentu (Blach, 2015). Pada proses normalisasi ini terdapat beberapa metode yang dapat dipakai, diantaranya adalah:

1. Metode Min-Max, yaitu metode normalisasi dengan melakukan transformasi linier terhadap data asli, dinyatakan sesuai persamaan (2.34):

$$DataBaru = \frac{(Data - DataMin) * (BaruMax - BaruMin)}{(DataMax - DataMin) + BaruMin} \quad (2.34)$$

Keterangan:

*DataBaru* : data hasil normalisasi

*Data* : data sebelum dinormalisasi

- DataMin* : nilai minimum data (lama) per kolom  
*DataMax* : nilai maksimum data (lama) per kolom  
*BaruMax* : batas maksimum data (baru) yang diinginkan  
*BaruMin* : batas minimum data (baru) yang diinginkan

2. Metode Z-score merupakan metode normalisasi yang berdasarkan nilai rata-rata (*mean*) dan deviasi standar (*standard deviation*) dari data, dinyatakan sesuai persamaan (2.35):

$$DataBaru = z = \frac{(Data - Mean)}{Std} \quad (2.35)$$

Keterangan:

- Mean* : nilai rata-rata dari data (lama) per kolom  
*Std* : nilai standard deviasi dari data (lama) per kolom

3. Metode Sigmoidal merupakan metode normalization melakukan normalisasi data secara nonlinier ke dalam range -1 s.d. 1 dengan menggunakan fungsi sigmoid, dinyatakan sesuai persamaan (2.36):

$$DataBaru = \frac{(1 - e^{-z})}{(1 + e^{-z})} \quad (2.36)$$

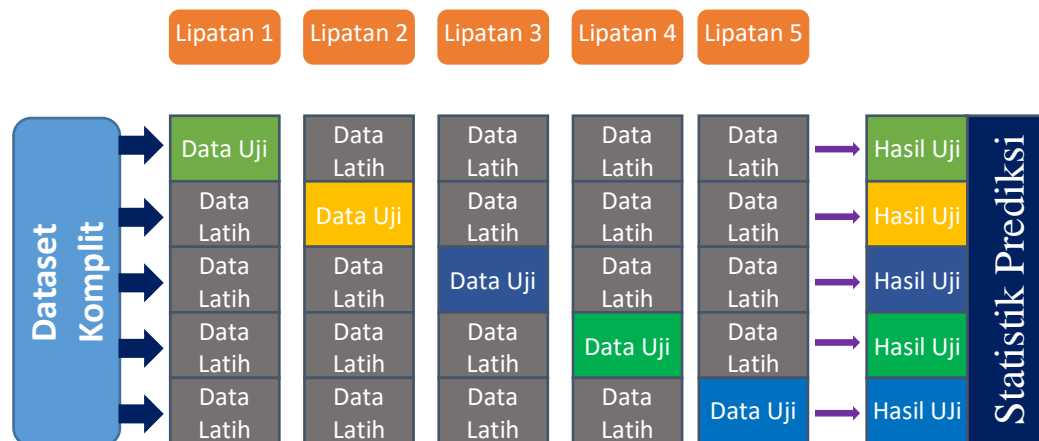
Keterangan:

- z* : hasil dari persamaan (2.35)  
*e* : nilai eksponensial (2,718281828)

## 2.6 Validasi Silang (*Cross Validation*)

Dalam *k-fold cross-validation* (Han et al., 2012), data awal secara acak dipartisi menjadi sebanyak *k* lipatan (*subset*) yang saling eksklusif, yang masing-masingnya berukuran kurang lebih sama. Pelatihan dan pengujian dilakukan sejumlah *k* kali (iterasi). Dalam iterasi *i*, partisi ke-*i* dicadangkan sebagai set tes, dan partisi yang tersisa secara kolektif digunakan untuk melatih model. Yaitu untuk contoh validasi silang 10 lipatan (*10-folds cross validation*), dalam iterasi pertama, subhimpunan lipatan ke-2 s.d. 10 secara kolektif berfungsi sebagai perangkat pelatihan untuk memperoleh model pertama, yang diuji pada lipatan ke-1; iterasi kedua dilatih pada subset lipatan ke-1, dan ke-3 s.d. 10 dan diuji pada

lipatan ke-2; dan seterusnya. Dalam *k-folds cross validation*, setiap sampel digunakan dalam jumlah yang sama untuk pelatihan dan sekali untuk pengujian. Untuk klasifikasi, perkiraan akurasi adalah jumlah keseluruhan klasifikasi yang benar dari  $k$  iterasi, dibagi dengan jumlah total iterasi. Untuk prediksi, perkiraan kesalahan dapat dihitung sebagai kerugian total dari  $k$  iterasi, dibagi dengan jumlah total iterasi. Secara umum, stratifikasi *10-folds cross-validation* direkomendasikan untuk memperkirakan akurasi (bahkan jika daya komputasi memungkinkan penggunaan lebih banyak lipatan) karena bias dan variansnya yang relatif rendah. Ilustrasi metode evaluasi validasi silang untuk jumlah lipatan sebesar 5 (lima), seperti yang disajikan pada Gambar 2.17.



Gambar 2.17 Contoh Validasi Silang 5 Lipatan (Helix, 2017)

Disarankan untuk menggunakan validasi silang 10 lipatan (*10-folds cross validation*), karena tes ekstensif yang telah dilakukan pada banyak dataset, dengan teknik pembelajaran yang berbeda, menunjukkan bahwa 10 adalah jumlah lipatan yang tepat untuk mendapatkan estimasi kesalahan terbaik, dan ada juga beberapa bukti teoritis yang mendukung hal ini. Meskipun argumen ini tidak berarti konklusif, dan perdebatan selalu ada tentang skema terbaik untuk evaluasi dalam pembelajaran mesin dan penambangan data. *10-fold cross-validation* telah menjadi metode standar dalam istilah praktis. Tes juga menunjukkan bahwa penggunaan stratifikasi dapat sedikit meningkatkan hasil. Dengan demikian teknik evaluasi standar di mana tersedia hanya data yang terbatas adalah *10-folds cross validation*. Harus diperhatikan bahwa baik stratifikasi maupun pembagian ke

dalam 10 lipatan harus dipenuhi syarat, yaitu tercukupi untuk membagi data menjadi 10 set yang kira-kira sama di mana berbagai nilai kelas dapat diwakili dalam kira-kira proporsi yang tepat. Evaluasi statistik bukanlah sebuah ilmu pasti. Selain itu, tidak dapat dipastikan manakah antara 10 lipatan, 5 lipatan atau 20 lipatan yang terbaik, karena cenderung memperoleh hasil yang hampir sama baiknya. (Witten & Frank, 2005)

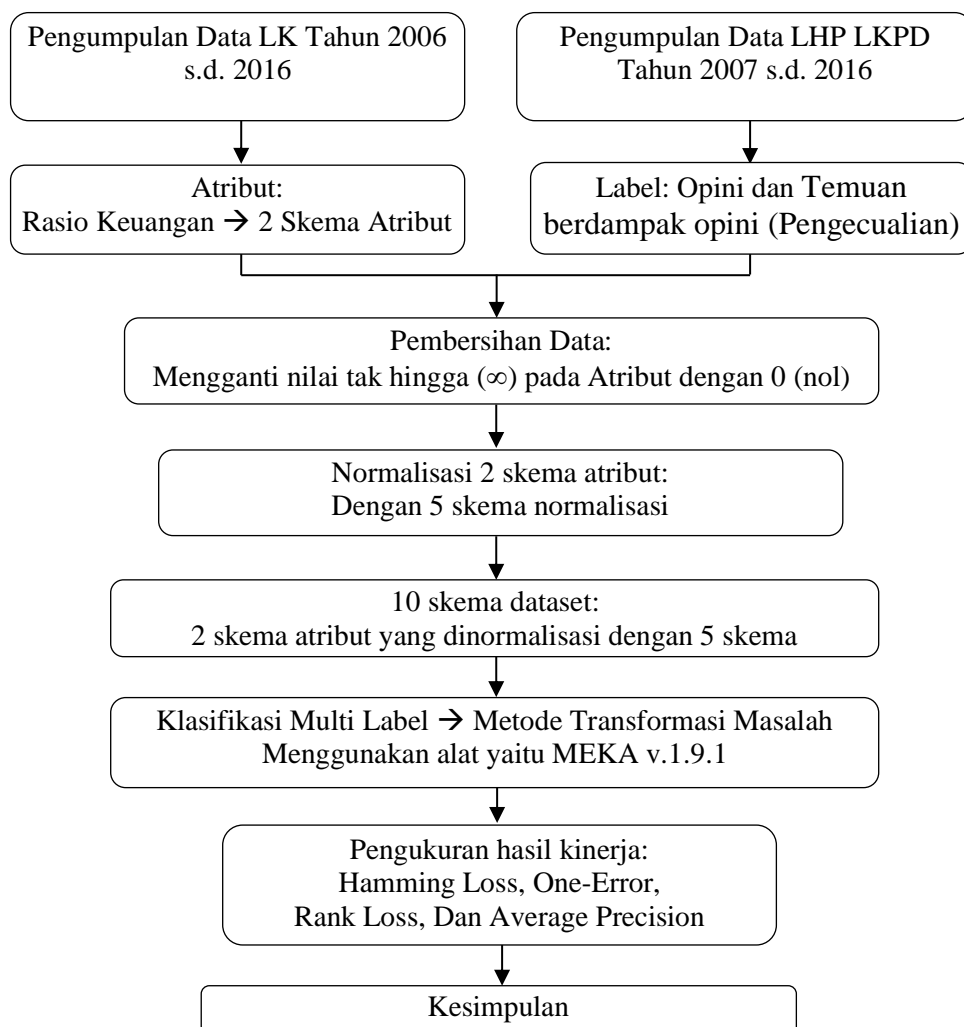
## BAB 3

### METODOLOGI PENELITIAN

Bab ini menjelaskan metode dan cara kerja yang dipakai dalam penelitian, sehingga dapat diketahui hasil kinerja penggunaan Metode Transformasi Masalah, yang merupakan salah satu pendekatan utama Klasifikasi Multi Label, dalam melakukan Prediksi Temuan.

#### 3.1 Alur Penelitian

Penelitian ini dimulai dari pengambilan data sampai dengan validasi hasil prediksi. Dengan alur penelitian seperti yang disajikan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

### **3.2 Pengumpulan Data**

Data yang digunakan dalam penelitian Prediksi Temuan ini terdiri atas atribut dan label sebanyak 150 data. Dengan atribut yang digunakan adalah rasio keuangan dari nilai akun-akun Laporan Keuangan yang diambil dari data Laporan Keuangan Tahun Anggaran 2006 s.d. 2016 di wilayah Provinsi Kalimantan Tengah. Data Laporan Keuangan Tahun Anggaran 2006 juga digunakan untuk memperoleh nilai akun-akun Laporan Keuangan “Tahun Sebelumnya” untuk data Laporan Keuangan Tahun Anggaran 2007, agar nilai rasio keuangan yang dijadikan atribut dapat diperoleh. Sedangkan untuk label diambil dari data Laporan Hasil Pemeriksaan atas Laporan Keuangan Pemerintah Daerah (LHP LKPD) Tahun Anggaran 2007 s.d. 2016 di wilayah Provinsi Kalimantan Tengah.

### **3.3 Atribut dan Label**

#### **3.3.1 Atribut**

Atribut yang digunakan dalam penelitian ini adalah rasio-rasio keuangan yang merupakan perbandingan dari nilai-nilai akun-akun Laporan Keuangan (LK) yang terdapat dalam Laporan Keuangan Pemerintah Daerah (LKPD) Tahun Anggaran (TA) 2006 s.d. 2016 pada 15 entitas wilayah pemeriksaan di wilayah Provinsi Kalimantan Tengah. Atribut tersebut diambil dari nilai akun-akun Laporan Keuangan (LK) yang terdapat pada Neraca, Laporan Realisasi Anggaran (LRA) dan Laporan Arus Kas (LAK).

Atribut tersebut diadopsi dari penelitian sebelumnya tentang Deteksi Opini Sektor Publik (Arianto et al., 2007a) atau Klasifikasi Data Keuangan Sektor Publik (Arianto, 2007b). Sehingga didapatkan atribut berupa rasio keuangan yang telah disesuaikan dengan akun-akun yang terdapat pada Laporan Keuangan Pemerintah Daerah (LKPD) di wilayah Provinsi Kalimantan Tengah (Kalteng). Daftar atribut yang digunakan beserta uraian rasio/perbandingan nilai akun Laporan Keuangan yang digunakan sebagai atribut serta nilai tersebut diperoleh dari komponen Laporan Keuangan yang mana.

Daftar atribut yang digunakan dalam penelitian Prediksi Temuan dapat dilihat pada Tabel 3.1.

Tabel 3.1 Daftar Atribut dalam penelitian Prediksi Temuan

Atribut	Uraian	Sumber
A1	Jumlah Kewajiban /Jumlah Kewajiban Tahun Sebelumnya	Neraca
A2	Jumlah Aset /Jumlah Aset Tahun Sebelumnya	Neraca
A3	SILPA /SILPA Tahun Sebelumnya	LRA
A4	PAD /PAD Tahun Sebelumnya	LRA
A5	Kas dan Setara Kas /Kas dan Setara Kas Tahun Sebelumnya	Neraca
A6	Piutang /Piutang Tahun Sebelumnya	Neraca
A7	Kas dan Setara Kas /PAD	Neraca - LRA
A8	Kas dan Setara Kas /Jumlah Aset	Neraca
A9	SILPA /Jumlah Aset	LRA - Neraca
A10	Jumlah Aset Lancar /Jumlah Aset	Neraca
A11	SILPA /PAD	LRA
A12	Piutang /PAD	Neraca - LRA
A13	PAD /Jumlah Aset	LRA - Neraca
A14	Jumlah Aset Lancar /Jumlah Kewajiban Jangka Pendek	Neraca
A15	PAD /Jumlah Aset Tetap	LRA - Neraca
A16	(Kas di Bendahara Pengeluaran + Kas di Bendahara Penerimaan) /Jumlah Aset	Neraca
A17	Kas dan Setara Kas /Jumlah Kewajiban Jangka Pendek	Neraca
A18	Jumlah Kewajiban /Jumlah Ekuitas Dana	Neraca
A19	Jumlah Kewajiban Jangka Panjang /Jumlah Aset	Neraca
A20	Jumlah Kewajiban /Jumlah Aset	Neraca
A21	Jumlah Aset /Jumlah Ekuitas Dana	Neraca
A22	Jumlah Kewajiban Jangka Panjang /Jumlah Ekuitas Dana	Neraca
A23	Jumlah Aset Tetap /Jumlah Aset	Neraca
A24	Jumlah Kas dan Setara Kas /Jumlah Aset Lancar	Neraca
A25	Piutang /Jumlah Aset	Neraca
A26	Jumlah Belanja /Jumlah Belanja Tahun Sebelumnya	LRA
A27	Jumlah Belanja Modal / Jumlah Aset Tetap - Jumlah Aset Tetap Tahun Sebelumnya	LRA - Neraca
A28	Belanja Barang Jasa / Persediaan - Persediaan Tahun Sebelumnya	LRA - Neraca
A29	Belanja Pegawai /Jumlah Belanja	LRA
A30	Jumlah Belanja Modal /Jumlah Belanja	LRA
A31	Belanja Hibah /Jumlah Belanja	LRA
A32	Belanja Bantuan Sosial /Jumlah Belanja	LRA
A33	PAD /Total Pendapatan Transfer	LRA

### 3.3.2 Label

Label yang digunakan dalam penelitian ini terdiri dari opini dan akun LK yang terdapat temuan berdampak opini (pengecualian) yang diambil dari LHP LKPD (Buku I) Tahun Anggaran 2007 s.d. 2016 pada 15 entitas di wilayah Provinsi Kalimantan Tengah. Label yang digunakan pada penelitian ini sebanyak 7 (tujuh) label (L1 s.d. L7). Dapat dilihat bahwa L1 merupakan label yang diambil dari opini atas LKPD, sedangkan L2 s.d. L7 merupakan data akun LK yang

menjadi pengecualian yang disebutkan pada LHP LKPD. Daftar label yang digunakan dapat dilihat pada Tabel 3.2.

Tabel 3.2 Daftar Label untuk penelitian Prediksi Temuan

Label	Uraian	Statistik data	Keterangan
Opini	0 : bila Opini selain WTP 1 : bila opini WTP	Jumlah 0 : 114 Jumlah 1 : 36	L1
Kas		Jumlah 0 : 121 Jumlah 1 : 29	L2
Piutang		Jumlah 0 : 144 Jumlah 1 : 6	-
Persediaan		Jumlah 0 : 133 Jumlah 1 : 17	-
Investasi Non Permanen		Jumlah 0 : 143 Jumlah 1 : 7	-
Penyertaan Modal	0 : bila tidak tercatat sebagai pengecualian (tidak mempengaruhi opini)	Jumlah 0 : 125 Jumlah 1 : 25	L3
Aset Tetap		Jumlah 0 : 75 Jumlah 1 : 75	L4
Aset Lainnya	1 : bila tercatat sebagai pengecualian (mempengaruhi opini)	Jumlah 0 : 141 Jumlah 1 : 9	-
Kewajiban Jangka Pendek		Jumlah 0 : 148 Jumlah 1 : 2	-
Pendapatan Asli Daerah (PAD)		Jumlah 0 : 129 Jumlah 1 : 21	L5
Belanja Operasi		Jumlah 0 : 85 Jumlah 1 : 65	L6
Belanja Modal		Jumlah 0 : 120 Jumlah 1 : 30	L7
Belanja Tidak Terduga		Jumlah 0 : 142 Jumlah 1 : 8	-
Belanja Transfer		Jumlah 0 : 148 Jumlah 1 : 2	-

Catatan: Label yang digunakan adalah dengan jumlah nilai 1 (satu) lebih dari 20 (dua puluh)

Contoh data untuk label yang diambil dari LHP LKPD dapat dilihat pada Gambar 3.2.

Menurut opini BPK, kecuali untuk dampak penyesuaian atas aset tetap, aset lain-lain, dan beban/akumulasi penyusutan tersebut, jika ada, yang mungkin perlu dilakukan jika Pemerintah Kota Palangka Raya melakukan pencatatan aset tetap yang memadai, memeriksa bukti-bukti aset tetap dan memperhitungan penambahan masa manfaat, laporan keuangan yang disebut di atas menyajikan secara wajar, dalam semua hal yang material, posisi keuangan Pemerintah Kota Palangka Raya tanggal 31 Desember 2015, dan realisasi anggaran, perubahan saldo anggaran lebih, operasional, arus kas, serta perubahan ekuitas untuk tahun yang berakhir pada tanggal-tanggal tersebut sesuai dengan Standar Akuntansi Pemerintahan.

Gambar 3.2 Contoh opini dan Pengecualian pada LHP LKPD (Buku I)



Dari contoh pada Gambar 3.2 diketahui bahwa opini Laporan Keuangan tersebut adalah Wajar Dengan Pengecualian (WDP) dan akun Laporan Keuangan yang menjadi pengecualian adalah akun Aset Tetap, Aset Lain-Lain, dan Beban/Akumulasi Penyusutan.

### 3.4 Pembersihan Data

Pada tahap praproses dilakukan pembersihan data, apabila terdapat nilai tak hingga ( $\infty$ ) pada rasio keuangan yang digunakan sebagai atribut, maka akan diganti dengan nol (0). Nilai tak hingga ( $\infty$ ) tersebut bisa ada dikarenakan terdapat akun yang menjadi pembagi pada rasio keuangan adalah bernilai nol, sehingga bilangan yang dibagi dengan nol (0) akan bernilai tak hingga ( $\infty$ ). Sehingga untuk mendapatkan data atribut yang baik, maka nilai tak hingga ( $\infty$ ) tersebut dirubah menjadi nol.

### 3.5 Skema Atribut

Penelitian ini dilakukan dengan membuat 2 (dua) skema atribut yang merupakan modifikasi dari sektor swasta (Ravisankar et al., 2011) dan atribut yang diambil dari sektor publik (Arianto et al., 2007a; Arianto, 2007b). Untuk pembagian skema atribut dapat dilihat pada Tabel 3.3.

Tabel 3.3 Skema Atribut

No.	Skema Atribut	Daftar Atribut	Keterangan
1.	F1	A1 s.d. A26	Modifikasi dari sektor swasta
2.	F2	A27 s.d. A33	Diambil dari sektor publik

### 3.6 Normalisasi

Setiap skema atribut (F1 dan F2) yang telah diperoleh akan dinormalisasi dengan 5 skema normalisasi. Setiap skema atribut dinormalisasi guna menghasilkan data dengan rentang (*range*) yang bisa diseragamkan, yaitu agar data yang digunakan tidak memiliki rentang (*range*) yang terlalu jauh. Dari 5 (lima) skema normalisasi, pada skema pertama yaitu normalisasi N1, data adalah sesuai awal atau tidak dilakukan normalisasi.

Skema normalisasi yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.4.

Tabel 3.4 Skema Normalisasi setiap dataset

No.	Uraian	Skema Normalisasi
1.	Tidak Dinormalisasi	N1
2.	Normalisasi Linear, skop 0 to 1	N2
3.	Normalisasi Non-Linear dengan z score	N3
4.	Normalisasi Sigmoidal Unipolar	N4
5.	Normalisasi dengan log	N5

Berikut persamaan yang digunakan untuk melakukan normalisasi yaitu persamaan (3.1) untuk skema normalisasi N2, persamaan (3.2) untuk skema normalisasi N3, persamaan (3.3) untuk skema normalisasi N4, dan persamaan (3.4) untuk skema normalisasi N5.

$$\hat{A}_{im} = \frac{A_{im} - \min(A_m)}{\max(A_m) - \min(A_m)} \quad (3.1)$$

$$\hat{A}_{im} = z = \frac{A_{im} - \bar{X}_m}{S_m} \quad (3.2)$$

$$\hat{A}_{im} = \frac{1}{1 + e^{-z}} \quad (3.3)$$

$$\hat{A}_{im} = \log(A_{im} - \min(A_m) + 1) \quad (3.4)$$

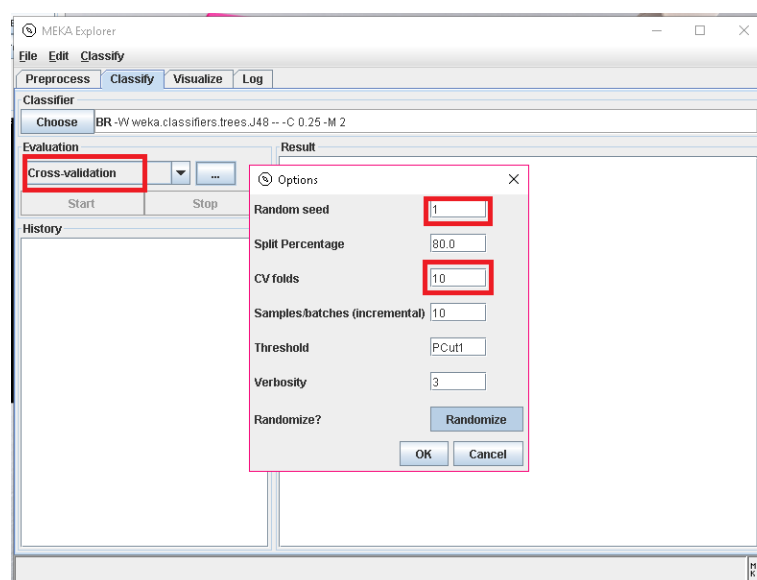
Keterangan:

- $A_{im}$  : Nilai awal atribut ke- $m$  data ke- $i$
- $\hat{A}_{im}$  : Hasil normalisasi atribut ke- $m$  data ke- $i$
- $\min(A_m)$  : Nilai terkecil pada setiap atribut ke- $m$
- $\max(A_m)$  : Nilai terbesar pada setiap atribut ke- $m$
- $\bar{A}_m$  : Nilai rerata pada setiap atribut ke- $m$
- $S_m$  : Nilai standar deviasi pada setiap atribut ke- $m$
- $i$  : Jumlah data
- $m$  : Jumlah atribut

### 3.7 Prediksi dengan Klasifikasi Multi Label

Penelitian ini menggunakan MEKA sebagai alat (*tool*) dalam melakukan Klasifikasi Multi Label (*multi label classification*) dengan pendekatan Metode

Transformasi Masalah (*Problem Transformation Methods*) untuk Prediksi Temuan. MEKA yang digunakan pada penelitian ini adalah versi 1.9.1 yang dirilis pada tanggal 12 April 2017. Dalam penelitian ini dilakukan pengaturan tambahan untuk evaluasi pada MEKA yaitu memilih metode *cross-validation* dengan pengaturan  $CVfold = 10$ ,  $Random\ seed = 1$  s.d. 10, sedangkan pengaturan yang lain dibiarkan sesuai setingan awal (*default*). Pengaturan untuk evaluasi pada MEKA tersebut seperti yang ditampilkan pada Gambar 3.3.



Gambar 3.3 Pengaturan Tambahan untuk evaluasi

Penelitian Prediksi Temuan ini menggunakan 3 (tiga) Metode Transformasi Masalah yang didukung oleh MEKA. Masing-masing Metode Transformasi Masalah tersebut dikombinasikan dengan 4 (empat) pengklasifikasi dasar yang juga didukung oleh MEKA. Kombinasi pengklasifikasi multi label dan pengklasifikasi dasar yang digunakan seperti yang disajikan pada Tabel 3.5.

Tabel 3.5 Pengklasifikasi Metode Transformasi Masalah yang digunakan

No.	Metode Transformasi Masalah	Pengklasifikasi Dasar
1.	Binary Relevance (BR)	a.J48
2.	Classifier Chain (CC)	b.SMO
3.	Label Combination (LC)	c. Random Forest (RF)
		d.Naive Bayes

### **3.8 Pengukuran Hasil Kinerja**

Hasil kinerja Metode Transformasi Masalah diukur dengan menampilkan nilai *Hamming Loss*, *One-Error*, *Rank Loss*, dan *Average Precision*. Pengukuran-pengukuran tersebut diambil dari hasil evaluasi yang dihasilkan oleh MEKA. Penilaian hasil kinerja akan dilakukan untuk setiap pengklasifikasi terhadap setiap skema dataset.

## BAB 4

### HASIL DAN PEMBAHASAN

Pada Bab ini akan diuraikan hasil dan pembahasan dari metode penelitian yang digunakan. Sehingga dapat diketahui hasil kinerja kombinasi pengklasifikasi Metode Transformasi Masalah (*Problem Transformation Methods*) pada Klasifikasi Multi Label (*multi label classification*) yang digunakan dalam melakukan Prediksi Temuan.

#### 4.1 Pembersihan Data

Dari data atribut yang diperoleh berdasarkan akun-akun pada Laporan Keuangan, terdapat beberapa atribut yang memiliki nilai tak hingga ( $\infty$ ). Nilai tak hingga ( $\infty$ ) tersebut dikarenakan ada pembagi pada komponen atribut yang merupakan rasio keuangan, yaitu perbandingan antara akun-akun pada Laporan Keuangan, yang bernilai 0 (nol) atau tidak ada nilai/isinya. Contoh pada Tabel 4.1, terdapat 22 data atribut A1 dengan nilai “Jumlah Kewajiban Tahun Sebelumnya” pada Laporan Keuangan yang bernilai 0 (nol) atau tidak ada nilai/isinya, sehingga nilai atribut A1, yaitu rasio antara Jumlah Kewajiban dengan Jumlah Kewajiban Tahun Sebelumnya, tersebut menjadi bernilai tak hingga ( $\infty$ ). Maka data atribut yang memiliki nilai tak hingga ( $\infty$ ) akan diganti dengan nilai 0 (nol), sehingga didapatkan data bersih yang layak digunakan untuk diujikan menggunakan Klasifikasi Multi Label khususnya Metode Transformasi Masalah untuk melakukan Prediksi Temuan. Pembersihan data ini dilakukan sebelum setiap skema atribut (F1 dan F2) dinormalisasi dengan 5 skema normalisasi. Daftar atribut yang dibersihkan karena terdapat nilai tak hingga ( $\infty$ ) sehingga diganti dengan 0 (nol) dapat dilihat pada Tabel 4.1.

Tabel 4.1 Atribut bernilai tak hingga ( $\infty$ ) yang diganti dengan nilai 0 (nol)

Atribut	Uraian	Jumlah Data
A1	Jumlah Kewajiban /Jumlah Kewajiban Tahun Sebelumnya	22
A6	Piutang /Piutang Tahun Sebelumnya	5
A14	Jumlah Aset Lancar /Jumlah Kewajiban Jangka Pendek	15
A17	Kas dan Setara Kas /Jumlah Kewajiban Jangka Pendek	15

## 4.2 Statistik Deskriptif Atribut

Setelah pembersihan data, sebelum melakukan normalisasi, dapat dilihat statistik deskriptif dari setiap atribut yang digunakan pada penelitian Prediksi Temuan seperti yang ditampilkan pada Tabel 4.2.

Tabel 4.2 Statistik Deskriptif Atribut

Atribut	Nilai Minimum	Nilai Maksimun	Nila Rerata	Nilai Standar Deviasi	Rentang
A1	0.0000	32392.6649	294.9142	2715.4893	32392.6649
A2	0.4502	2.0145	1.1511	0.1848	1.5643
A3	0.1639	9.2050	1.2777	1.1047	9.0410
A4	0.6101	3.8752	1.2216	0.3476	3.2651
A5	0.1702	8.3469	1.2513	0.9642	8.1767
A6	0.0000	1753.0186	16.5189	143.4472	1753.0186
A7	0.0463	14.4550	4.1174	3.0498	14.4088
A8	0.0056	0.1764	0.0642	0.0346	0.1708
A9	0.0044	0.1764	0.0643	0.0352	0.1720
A10	0.0068	0.1827	0.0729	0.0349	0.1759
A11	0.0335	14.1195	4.1049	3.0534	14.0860
A12	0.0000	0.7185	0.1700	0.1646	0.7185
A13	0.0049	0.1415	0.0254	0.0261	0.1365
A14	0.0000	1895046.9780	26552.9242	196009.6459	1895046.9780
A15	0.0051	0.1695	0.0288	0.0303	0.1644
A16	0.0000	0.0105	0.0009	0.0019	0.0105
A17	0.0000	1711783.2640	22894.1080	170415.4469	1711783.2640
A18	0.0000	0.0516	0.0050	0.0084	0.0516
A19	0.0000	0.0211	0.0011	0.0035	0.0211
A20	0.0000	0.0490	0.0049	0.0081	0.0490
A21	1.0000	1.0516	1.0050	0.0084	0.0516
A22	0.0000	0.0216	0.0011	0.0036	0.0216
A23	0.7631	0.9884	0.8936	0.0422	0.2253
A24	0.1644	0.9951	0.8623	0.1359	0.8306
A25	0.0000	0.0286	0.0038	0.0046	0.0286
A26	0.8670	1.7393	1.1326	0.1302	0.8723
A27	0.0931	11.3770	1.1485	1.1471	11.2839
A28	3.6165	8292.5112	323.2753	845.2571	8288.8947
A29	0.1476	0.6461	0.3823	0.1071	0.4985
A30	0.1574	0.6191	0.3476	0.0991	0.4617
A31	0.0000	0.2646	0.0340	0.0401	0.2646
A32	0.0000	0.1157	0.0149	0.0164	0.1157
A33	0.0192	0.8026	0.0845	0.1351	0.7834

### 4.3 Skema Dataset

Dataset yang digunakan pada penelitian ini terdiri atas 10 skema, yang diperoleh dari 2 skema atribut yang dinormalisasi dengan 5 skema normalisasi. 10 skema dataset terdiri atas atribut dan label. Skema dataset yang digunakan dapat dilihat pada Tabel 4.3.

Tabel 4.3 Skema dataset

No.	Skema Dataset	Keterangan
1.	F1_N1	Skema Atribut F1 dinormalisasi dengan skema normalisasi N1
2.	F1_N2	Skema Atribut F1 dinormalisasi dengan skema normalisasi N2
3.	F1_N3	Skema Atribut F1 dinormalisasi dengan skema normalisasi N3
4.	F1_N4	Skema Atribut F1 dinormalisasi dengan skema normalisasi N4
5.	F1_N5	Skema Atribut F1 dinormalisasi dengan skema normalisasi N5
6.	F2_N1	Skema Atribut F2 dinormalisasi dengan skema normalisasi N1
7.	F2_N2	Skema Atribut F2 dinormalisasi dengan skema normalisasi N2
8.	F2_N3	Skema Atribut F2 dinormalisasi dengan skema normalisasi N3
9.	F2_N4	Skema Atribut F2 dinormalisasi dengan skema normalisasi N4
10.	F2_N5	Skema Atribut F2 dinormalisasi dengan skema normalisasi N5

### 4.4 Klasifikasi menggunakan MEKA

Sebanyak 10 (sepuluh) skema dataset yang diperoleh dari 2 (dua) skema atribut (F1 dan F2) yang dinormalisasi dengan 5 (lima) skema normalisasi (N1 s.d. N5) lalu diujikan menggunakan MEKA versi 1.9.1. Setiap skema dataset diujikan dengan 12 (dua belas) kombinasi pengklasifikasi, yaitu kombinasi dari 3 (tiga) pengklasifikasi multi label Metode Transformasi Masalah yang masing-masingnya dikombinasikan dengan 4 (empat) pengklasifikasi dasar, yang semua kombinasi pengklasifikasi tersebut didukung oleh MEKA. Semua pengklasifikasi tersebut diatur sesuai setingan awal dari MEKA, atau sesuai *default*. Untuk evaluasi hasil pengujian performa kombinasi pengklasifikasi digunakan metode validasi silang 10 lipatan (*10-fold cross validation*) dengan nilai benih acak (*Random Seed*) 1 s.d. 10. Untuk mengetahui hasil pengujian (evaluasi matriks) dari setiap kombinasi pengklasifikasi yang diujikan pada setiap skema dataset, maka akan diambil nilai hasil evaluasi yaitu *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision*.

## 4.5 Hasil Matriks Evaluasi

Berikut adalah hasil evaluasi matriks pengukuran kinerja pengklasifikasi yang diujikan pada setiap skema dataset:

### 4.5.1 Hasil Hamming Loss

Dari nilai hasil pengukuran *Hamming Loss* untuk setiap dataset diketahui bahwa nilai *Hamming Loss* terbaik atau nilai terkecil adalah sebesar 0,194. Nilai *Hamming Loss* terbaik tersebut diperoleh dari pengujian dataset F2\_N3 menggunakan kombinasi pengklasifikasi Metode Transformasi Masalah *Binary Relevance* dengan pengklasifikasi dasar *Random Forest* (BR-RF). Hasil pengukuran metrik evaluasi *Hamming Loss* dapat dilihat pada Tabel 4.3.

Tabel 4.3 Hasil *Hamming Loss* dari setiap dataset

Hamming Loss		BR				CC				LC			
No.	Dataset	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	0.259	0.234	0.233	0.491	0.289	0.248	0.237	0.475	0.300	0.285	0.271	0.317
2	F1_N2	0.260	0.235	0.236	0.491	0.286	0.248	0.236	0.475	0.288	0.286	0.273	0.317
3	F1_N3	0.256	0.234	0.235	0.490	0.286	0.248	0.235	0.475	0.295	0.286	0.270	0.318
4	F1_N4	0.255	0.219	0.235	0.441	0.287	<b>0.230</b>	0.236	0.428	0.292	0.262	0.274	0.305
5	F1_N5	0.261	0.240	0.231	0.405	0.288	0.246	0.236	0.397	0.309	0.281	0.273	0.305
6	F2_N1	<b>0.215</b>	0.231	0.196	0.313	<b>0.217</b>	0.271	<b>0.196</b>	0.340	0.249	0.312	<b>0.218</b>	0.265
7	F2_N2	0.216	0.231	0.197	0.313	<b>0.217</b>	0.271	0.197	0.341	<b>0.248</b>	0.313	0.222	0.266
8	F2_N3	0.216	0.231	<b>0.194</b>	0.313	<b>0.217</b>	0.271	<b>0.196</b>	0.340	0.249	0.313	0.223	0.261
9	F2_N4	0.216	<b>0.204</b>	0.199	<b>0.255</b>	<b>0.217</b>	0.231	0.198	<b>0.274</b>	0.249	<b>0.243</b>	0.224	<b>0.242</b>
10	F2_N5	<b>0.215</b>	0.227	0.195	0.268	<b>0.217</b>	0.268	0.197	0.298	0.251	0.315	0.219	0.254

### 4.5.2 Hasil One-Error

Dari nilai hasil pengukuran *One-Error* untuk setiap dataset diketahui bahwa nilai *One-Error* terbaik atau nilai terkecil adalah sebesar 0,270. Nilai *One-Error* terbaik tersebut diperoleh dari pengujian dataset F2\_N5 menggunakan kombinasi pengklasifikasi Metode Transformasi Masalah *Binary Relevance* dengan pengklasifikasi dasar *Random Forest* (BR-RF).



Hasil pengukuran metrik evaluasi *One-Error* atas pengujian pada setiap skema dataset dapat dilihat pada Tabel 4.4.

Tabel 4.4 Hasil *One-Error* dari setiap dataset

One-Error		BR				CC				LC			
No.	Dataset	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	0.519	0.425	0.379	0.664	0.545	0.427	0.427	0.634	0.595	0.548	0.491	0.617
2	F1_N2	0.532	0.429	0.374	0.659	0.539	0.429	0.429	0.639	0.567	0.551	0.503	0.613
3	F1_N3	0.539	0.425	0.381	0.667	0.551	0.429	0.417	0.637	0.583	0.553	0.496	0.613
4	F1_N4	0.533	0.369	0.385	0.593	0.552	0.377	0.426	0.594	0.577	0.443	0.498	0.594
5	F1_N5	0.523	0.450	0.375	0.565	0.537	0.429	0.424	0.625	0.615	0.534	0.501	0.583
6	F2_N1	<b>0.377</b>	0.395	0.271	0.506	0.327	0.507	0.327	0.617	<b>0.462</b>	0.651	<b>0.373</b>	0.557
7	F2_N2	0.379	0.396	0.278	0.508	0.327	0.508	0.332	0.617	0.463	0.653	0.389	0.555
8	F2_N3	0.379	0.397	0.273	0.512	<b>0.325</b>	0.507	<b>0.321</b>	0.621	0.464	0.654	0.381	0.543
9	F2_N4	0.379	<b>0.364</b>	0.279	<b>0.354</b>	<b>0.325</b>	<b>0.361</b>	0.328	<b>0.468</b>	0.464	<b>0.407</b>	0.386	<b>0.491</b>
10	F2_N5	<b>0.377</b>	0.380	<b>0.270</b>	0.402	0.329	0.495	0.325	0.507	0.467	0.661	<b>0.373</b>	0.528

#### 4.5.3 Hasil Rank Loss

Dari nilai hasil pengukuran *Rank Loss* untuk setiap dataset diketahui bahwa nilai *Rank Loss* terbaik atau nilai terkecil adalah sebesar 0,166. Nilai *Rank Loss* terbaik tersebut diperoleh dari pengujian dataset F2\_N2 dan F2\_N3 menggunakan kombinasi pengklasifikasi Metode Transformasi Masalah *Binary Relevance* dengan pengklasifikasi dasar *Random Forest* (BR-RF). Hasil pengukuran metrik evaluasi *Rank Loss* dapat dilihat pada Tabel 4.5.

Tabel 4.5 Hasil *Rank Loss* dari setiap dataset

Rank Loss		BR				CC				LC			
No.	Dataset	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	0.336	0.391	0.216	0.442	0.403	0.399	0.366	0.464	0.398	0.438	0.383	0.429
2	F1_N2	0.331	0.394	0.218	0.443	0.410	0.399	0.360	0.462	0.379	0.440	0.385	0.422
3	F1_N3	0.328	0.392	0.216	0.443	0.406	0.400	0.360	0.462	0.393	0.441	0.378	0.426
4	F1_N4	0.326	0.320	0.219	0.407	0.408	<b>0.335</b>	0.362	0.439	0.390	0.381	0.379	0.403
5	F1_N5	0.333	0.401	0.219	0.352	0.397	0.395	0.362	0.415	0.411	0.432	0.381	0.398
6	F2_N1	<b>0.253</b>	0.441	0.168	0.289	<b>0.277</b>	0.462	0.250	0.345	<b>0.298</b>	0.499	<b>0.269</b>	0.329
7	F2_N2	<b>0.253</b>	0.443	<b>0.166</b>	0.288	<b>0.277</b>	0.463	0.253	0.345	<b>0.298</b>	0.499	0.272	0.331
8	F2_N3	<b>0.253</b>	0.442	<b>0.166</b>	0.289	<b>0.277</b>	0.462	<b>0.249</b>	0.347	0.299	0.501	0.275	0.323
9	F2_N4	<b>0.253</b>	<b>0.311</b>	0.168	<b>0.217</b>	<b>0.277</b>	0.337	0.251	<b>0.312</b>	0.299	<b>0.340</b>	0.279	<b>0.263</b>
10	F2_N5	<b>0.253</b>	0.431	0.167	0.236	<b>0.277</b>	0.453	0.249	0.321	0.300	0.506	0.271	0.273

#### 4.5.4 Hasil Average Precision

Dari nilai hasil pengukuran *Average Precision* untuk setiap dataset diketahui bahwa nilai *Average Precision* terbaik atau nilai terbesar adalah sebesar 0,787. Nilai *Average Precision* terbaik tersebut diperoleh dari pengujian dataset F2\_N3 dan F2\_N3 menggunakan kombinasi pengklasifikasi Metode Transformasi Masalah *Binary Relevance* dengan pengklasifikasi dasar *Random Forest* (BR-RF). Hasil pengukuran metrik evaluasi *Average Precision* disajikan pada Tabel 4.6.

Tabel 4.6 Hasil *Average Precision* pada setiap dataset

Avg Precision		BR				CC				LC			
No.	Dataset	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	0.619	0.492	0.719	0.521	0.504	0.550	0.547	0.400	0.502	0.553	0.562	0.463
2	F1_N2	0.614	0.490	0.718	0.525	0.505	0.550	0.551	0.404	0.523	0.552	0.559	0.469
3	F1_N3	0.614	0.492	0.716	0.521	0.504	0.550	0.551	0.402	0.507	0.552	0.561	0.466
4	F1_N4	0.614	0.568	0.712	0.561	0.504	0.594	0.552	0.439	0.518	0.572	0.555	0.488
5	F1_N5	0.616	0.484	0.715	0.598	0.509	0.556	0.551	0.465	0.489	0.557	0.558	0.501
6	F2_N1	<b>0.714</b>	0.486	0.786	0.645	<b>0.619</b>	0.495	0.652	0.483	0.579	0.519	0.634	0.524
7	F2_N2	0.713	0.485	0.784	0.645	<b>0.619</b>	0.495	0.649	0.483	<b>0.581</b>	0.519	0.630	0.523
8	F2_N3	0.713	0.486	<b>0.787</b>	0.643	<b>0.619</b>	0.495	<b>0.652</b>	0.485	0.580	0.517	0.630	0.530
9	F2_N4	<b>0.714</b>	<b>0.596</b>	0.783	<b>0.736</b>	<b>0.619</b>	<b>0.601</b>	0.651	<b>0.573</b>	0.580	<b>0.618</b>	0.627	<b>0.588</b>
10	F2_N5	<b>0.714</b>	0.495	0.785	0.703	<b>0.619</b>	0.501	0.650	0.532	0.577	0.513	<b>0.636</b>	0.574

#### 4.5.5 Rekapitulasi Hasil Evaluasi

Hasil rekapitulasi menampilkan nilai terbaik setiap pengukuran evaluasi kinerja (*Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision*) terhadap pengujian kombinasi pengklasifikasi pada setiap dataset yang digunakan dalam penelitian ini. Dari rekapitulasi tersebut dapat diketahui bahwa kombinasi pengklasifikasi yang memperoleh hasil terbaik pada setiap pengukuran *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision* adalah pengklasifikasi multi label *Binary Relevance* dikombinasikan dengan pengklasifikasi dasar *Random Forest* (BR-RF).

Hasil rekapitulasi pengukuran metrik evaluasi *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision* dapat dilihat pada Tabel 4.7.

Tabel 4.7 Hasil evaluasi untuk setiap skema dataset

No.	Dataset	Evaluation Metrics			
		<i>Hamming Loss</i>	<i>One-Error</i>	<i>Rank Loss</i>	<i>Average Precision</i>
1.	F1_N1	0.233 (BR-RF)	0.379 (BR-RF)	0.216 (BR-RF)	0.719 (BR-RF)
2.	F1_N2	0.235 (BR-SMO)	0.374 (BR-RF)	0.218 (BR-RF)	0.718 (BR-RF)
3.	F1_N3	0.234 (BR-SMO)	0.381 (BR-RF)	0.216 (BR-RF)	0.716 (BR-RF)
4.	F1_N4	0.219 (BR-SMO)	0.369 (BR-SMO)	0.219 (BR-RF)	0.712 (BR-RF)
5.	F1_N5	0.231 (BR-RF)	0.375 (BR-RF)	0.219 (BR-RF)	0.715 (BR-RF)
6.	F2_N1	0.196 (BR-RF)	0.271 (BR-RF)	0.168 (BR-RF)	0.786 (BR-RF)
7.	F2_N2	0.197 (BR-RF)	0.278 (BR-RF)	<b>0.166</b> <b>(BR-RF)</b>	0.784 (BR-RF)
8.	F2_N3	<b>0.194</b> <b>(BR-RF)</b>	0.273 (BR-RF)	<b>0.166</b> <b>(BR-RF)</b>	<b>0.787</b> <b>(BR-RF)</b>
9.	F2_N4	0.198 (CC-RF)	0.279 (BR-RF)	0.168 (BR-RF)	0.783 (BR-RF)
10.	F2_N5	0.195 (BR-RF)	<b>0.270</b> <b>(BR-RF)</b>	0.167 (BR-RF)	0.785 (BR-RF)

#### 4.6 Analisa Hasil

Dari hasil pengukuran kinerja setiap kombinasi pengklasifikasi pada setiap skema dataset, dapat diketahui hasil pengujian pada setiap skema atribut F1 dan F2 yang dinormalisasi dengan skema normalisasi N1, N2, N3, N4, dan N5. Dari penggunaan skema atribut F2, diketahui skema atribut F2 yang dinormalisasi dengan skema normalisasi N2, N3 dan N5 yang menghasilkan nilai terbaik untuk setiap pengukuran. Skema dataset F2\_N2, memberikan hasil terbaik pada pengukuran *Rank Loss*. Skema dataset F2\_N5 memberikan hasil terbaik pada pengukuran *One-Error*. Sedangkan skema dataset F2\_N3 memberikan hasil terbaik pada pengukuran *Hamming Loss*, *Rank Loss*, dan *Average Precision*.

Penggunaan atribut/fitur pada penelitian ini yang sama dengan penelitian sebelumnya terkait Deteksi Opini Sektor Publik (Arianto et al., 2017a) atau Klasifikasi Data Keuangan Sektor Publik (Arianto, 2017b), juga menunjukkan bahwa penggunaan atribut yang diambil dari sektor publik (F2) memberikan hasil

yang lebih baik dibandingkan penggunaan atribut yang merupakan modifikasi dari sektor swasta (F1).

Dari pengujian 12 (dua belas) kombinasi pengklasifikasi pada 10 (sepuluh) skema dataset, diketahui bahwa pengklasifikasi multi label *Binary Relevance* (BR) yang dikombinasikan dengan pengklasifikasi dasar *Random Forest* (RF) memberikan hasil terbaik pada setiap pengukuran kinerja/performa pengklasifikasi.

Dari data yang diperoleh terkait label yaitu akun LK yang menjadi pengecualian (L2 s.d. L7), maka akun LK yang paling sering muncul sebagai pengecualian dapat diurutkan. Daftar akun LK yang menjadi pengecualian berdasarkan data yang diperoleh, setelah diurutkan disajikan pada Tabel 4.8.

Tabel 4.8 Statistik Deskriptif Label yang diurutkan

No.	Kode Label	Uraian Label	Statistik Label
1.	L4	Aset Tetap	75
2.	L6	Belanja Operasi	65
3.	L7	Belanja Modal	30
4.	L2	Kas	29
5.	L3	Penyertaan Modal	25
6.	L5	Pendapatan Asli Daerah (PAD)	21

Dari data Tabel 4.8 diketahui bahwa akun LK yang sering muncul atau menjadi pengecualian dari data LHP LKPD wilayah Provinsi Kalimantan Tengah TA 2007 s.d. 2016 sesuai urutan adalah Aset Tetap, Belanja Operasi, Belanja Modal, Kas, Penyertaan Modal, dan Pendapatan Asli Daerah (PAD). Sehingga dapat dijadikan sebagai informasi awal bagi pemeriksa dalam melakukan tugas pemeriksaan LK yaitu untuk menentukan akun LK yang mana yang didahulukan, menjadi prioritas, untuk dijadikan sampling pemeriksaan Laporan Keuangan.

## **BAB 5**

### **PENUTUP**

#### **5.1 Kesimpulan**

Penelitian Prediksi Temuan ini menggunakan Metode Transformasi Masalah yang merupakan salah satu pendekatan dalam melakukan Klasifikasi Multi Label. Hasil dari penelitian ini diharapkan dapat menjadi alat (*tool*) yang mampu membantu pemeriksa BPK RI dalam melakukan pemeriksaan Laporan Keuangan, khususnya dalam menentukan akun-akun Laporan Keuangan yang menjadi prioritas untuk disampling dalam pelaksanaan tugas pemeriksaan karena diprediksi terdapat temuan-temuan signifikan yang dapat berpengaruh/berdampak pada penentuan opini.

Data yang digunakan adalah hasil pemeriksaan Laporan Keuangan di wilayah Provinsi Kalimantan Tengah Tahun Anggaran 2007 s.d. 2016. Sehingga didapat 150 data yang terdiri atas 7 (tujuh) label dan 33 (tiga puluh tiga) atribut yang dibagi menjadi 2 (dua) skema atribut (F1 dan F2). Masing-masing skema atribut dinormalisasi menggunakan 5 (lima) skema normalisasi, sehingga diperoleh 10 skema dataset yang digunakan dalam penelitian ini.

Penelitian ini menggunakan aplikasi MEKA versi 1.9.1 sebagai alat (*tool*) dalam melakukan klasifikasi multi label. Pada pengujian, setiap dataset diuji menggunakan 12 (dua belas) kombinasi pengklasifikasi yaitu 3 (tiga) pengklasifikasi Metode Transformasi Masalah antara lain *Binary Relevance* (BR), *Classifier Chains* (CC), dan *Label Combination* (LC) yang dikombinasikan dengan 4 (empat) pengklasifikasi dasar yang didukung oleh MEKA, antara lain J48, SMO, *Random Forest* (RF), dan *Naive Bayes* (NB).

Untuk penilaian kinerja/performa setiap kombinasi pengklasifikasi pada pengujian setiap skema dataset, dilakukan pengukuran metrik evaluasi antara lain: *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision*. Untuk evaluasi pada MEKA, digunakan metode validasi silang 10 lipatan (*10-fold cross validation*) dengan nilai *Random Seed* 1 s.d. 10. Dari hasil evaluasi kinerja pada 12 kombinasi pengklasifikasi yang diujikan pada 10 skema dataset, diketahui

bahwa nilai terbaik pengukuran *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision* diperoleh dari pengujian menggunakan kombinasi pengklasifikasi *Binary Relevance* dengan *Random Forest* (BR-RF).

Hasil pengujian setiap skema dataset, diketahui bahwa penggunaan skema atribut F2 yang dinormalisasi dengan skema normalisasi N2, N3 dan N5 memberikan hasil nilai terbaik pada beberapa pengukuran. Sedangkan skema dataset F2\_N3 memberikan hasil terbaik hampir pada semua pengukuran kinerja pengklasifikasi kecuali pada pengukuran metrik evaluasi *One-Error*. Hasil pengujian juga menunjukkan bahwa penggunaan atribut skema F2 (A27 s.d. A33) yang diambil dari sektor publik, memberikan hasil yang lebih baik daripada penggunaan atribut skema F1 (A1 s.d. A26), yang merupakan modifikasi dari sektor swasta, pada semua pengukuran kinerja kombinasi pengklasifikasi.

Dari pengukuran *Hamming Loss*, *One-Error*, *Rank Loss* dan *Average Precision* diketahui bahwa penggunaan Klasifikasi Multi Label khususnya Metode Transformasi Masalah memberikan hasil yang cukup baik dalam melakukan Prediksi Temuan. Dari nilai metrik evaluasi *Hamming Loss*, diketahui bahwa terdapat kesalahan prediksi label sebesar 0,194 (19,4%).

Dari statistik deskriptif data label terkait akun LK yang menjadi pengecualian (L2 s.d. L7), dapat diurutkan akun LK yang paling sering muncul sebagai pengecualian dari data LHP LKPD wilayah Provinsi Kalimantan Tengah TA 2007 s.d. 2016, sesuai urutan adalah Aset Tetap, Belanja Operasi, Belanja Modal, Kas, Penyertaan Modal, dan Pendapatan Asli Daerah (PAD). Sehingga hal tersebut dapat dijadikan sebagai informasi awal bagi pemeriksaaan dalam menentukan akun LK yang mana yang terlebih dahulu menjadi prioritas untuk dijadikan sampling dalam pemeriksaaan Laporan Keuangan.

## 5.2 Saran

Untuk penelitian selanjutnya yang masih berkaitan dengan pemeriksaaan Laporan Keuangan dan klasifikasi multi label, dapat dilakukan penambahan atau pengembangan diantaranya:

1. Penggunaan pengklasifikasi Metode Transformasi Masalah lain yang merupakan pengembangan dari *Binary Relevance* (BR), *Classifier Chains* (CC), dan *Label Combination* (LC);
2. Penggunaan pengklasifikasi Metode Adaptasi Algoritma yang juga merupakan salah satu pendekatan untuk klasifikasi multi label;
3. Penggunaan atribut selain yang berupa rasio keuangan (Laporan Keuangan) dan data lain yang masih berkaitan dengan pemeriksaan Laporan Keuangan;
4. Penggunaan data dengan jumlah yang lebih banyak, pada penelitian ini menggunakan 150 data yang diperoleh dari data hasil pemeriksaan TA 2007 s.d. 2016 (10 tahun) pada 15 entitas pemeriksaan di wilayah Provinsi Kalimantan Tengah.

*Halaman ini sengaja dikosongkan*



## DAFTAR PUSTAKA

- Seyed Mojtaba Saif, Mehdi Sarikhani, and Fahime Ebrahimi (2012), "Finding rules for audit opinions prediction through data mining methods," *European Online Journal of Natural and Social Sciences*, vol. 1, no.2, hal. 28-36.
- Alpaslan Yaşar, Emre Yakut, and M. Murat Gutnu (2015), "Predicting Qualified Audit Opinions Using Financial Ratios: Evidence from the Istanbul Stock Exchange," *International Journal of Business and Social Science*, vol. 6, no. 8, hal. 57-67.
- Omid Pourheydari, Hossein Nezamabadi-pour, and Zeinab Aazami (2012), "Identifying qualified audit opinions by artificial neural networks," *African Journal of Business Management*, vol. 6, no. 44, hal. 11077-11087.
- Daniel Zdolšek, Timotej Jagrič, and Marjan Odar (2015), "Identification of auditor's report qualifications: an empirical analysis for Slovenia," *Economic Research-Ekonomska Istraživanja*, vol. 28, no. 1, hal. 994-1005.
- Ahmad Dwi Arianto, Achmad Affandi, and Supeno Mardi Susiki Nugroho (2017a), "Opinion Detection of Public Sector Financial Statements Using K-Nearest Neighbors," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 4, hal. 1-5.
- Ahmad Dwi Arianto (2017b), *Klasifikasi Data Keuangan Sektor Publik Untuk Penentuan Sampel Pemeriksaan Menggunakan K-Nearest Neighbors*, Tesis Master, Institut Teknologi Sepuluh Nopember, Surabaya.
- Michael C. Ehrhardt, Eugene F. Brigham (2011), *Financial Management: Theory and Practice*, 13th edition, South-Western Cengage Learning, Mason, Ohio, USA.
- Ziad Abdallah, Ali El-Zaart, and Mohamad Oueidat (2015), "Comparison of Multilabel Problem Transformation Methods for Text Mining," *International Conference on Digital Information and Communication Technology and its Applications (DICTAP)*, vol. 5, hal. 115-118.

- Kavitha C.R., and Mahalekshmi T. (2016), “A Comparison of Multi-Label Classification Methods Using MEKA on Benchmark Datasets,” *International Journal of Research in Engineering and Technology (IJRET)*, vol. 05, no. 9, hal. 330-335.
- Francisco Herrera, Francisco Charte, Antonio J. Rivera and María J. del Jesus (2016), *Multilabel Classification Problem Analysis, Metrics and Techniques*, Springer, Granada, Spain, Jaén, Spain.
- Grigorios Tsoumakas, and Ioannis Katakis (2007), “Multi-label classification: an overview”, *International Journal of Data Warehouse and Mining*, vol. 3, no. 3, hal. 1-13.
- Matthew R. Boutella, Jiebo Luo, Xipeng Shen, and Christopher M. Brown (2004), “Learning multi-label scene classification”, *Pattern Recognition*, vol. 37, no. 9, hal. 1757-1771.
- Jesse Read (2010), *Scalable Multi-label Classification*, Thesis Ph.D., The University of Waikato, Hamilton, New Zealand.
- Eva Gibaja, and Sebastian Ventura (2010), “A Tutorial on Multi-Label Learning”, *ACM Computing Surveys*, vol. 9, no. 4, article 39, hal. 1-39.
- Jesse Read, Bernhard Pfahringer, Geoffrey Holmes, and Eibe Frank (2009), “Classifier chains for multi-label classification”, *In Proceeding(s) of the ECML/PKDD*, vol. 5782, hal. 254-269.
- Jesse Read, Luca Martino, Pablo M. Olmos, and David Luengo (2015), “Scalable multi-output label prediction: From classifier chains to classifier trellises”, *Pattern Recognition*, vol. 48, hal. 2096-2109.
- Min-Ling Zhang, and Zhi-Hua Zhou (2007), “ML-KNN: A lazy learning approach to multi-label learning,” *Pattern Recognition*, vol. 40, hal. 2038-2048.
- Everton Alvares Cherman, Maria Carolina Monard, and Jean Metz (2011), “Multi-label Problem Transformation Methods: a Case Study,” *CLEI Electronic Journal*, vol. 14, no. 1, hal. 1-10.

- Saed Sayad, diakses 24 Mei 2017, [http://www.saedsayad.com/decision\\_tree.htm](http://www.saedsayad.com/decision_tree.htm), Dr. Saed Sayad (2010-2018a) *Decision Tree - Classification*, [Online].
- Saed Sayad, Diakses 24 Mei 2017, [http://www.saedsayad.com/support\\_vector\\_machine.htm](http://www.saedsayad.com/support_vector_machine.htm), Dr. Saed Sayad (2010-2018b), *Support Vector Machine - Classification (SVM)*, [Online].
- Jiawei Han, and Micheline Kamber (2006), *Data Mining Concepts and Techniques*, 2nd edition, Morgan Kaufmann, San Francisco, USA.
- Saed Sayad, diakses 24 Mei 2017, [http://www.saedsayad.com/naive\\_bayesian.htm](http://www.saedsayad.com/naive_bayesian.htm), Dr. Saed Sayad (2010-2018c), *Naive Bayesian*, [Online].
- Jiawei Han, Micheline Kamber, and Jian Pei (2012), *Data Mining Concepts and Techniques*, 3rd edition, Morgan Kaufmann, Waltham, USA.
- Adele Cutler, David Richard Cutler, and John R. Stevens (2012), "Random Forests," dalam *Ensemble Machine Learning*, Cha Zhang, Yunqian Ma, eds., Springer, Boston, hal. 157-175
- Jesse Read, Peter Reutemann, Bernhard Pfahringer, and Geo Holmes (2016), "Meka: A Multi-label/Multi-target Extension to Weka," *Journal of Machine Learning Research*, vol. 17, hal. 1-5.
- Sourceforge, diakses 03 Desember 2017, <http://meka.sourceforge.net/>, Jesse Read (2012-2017), *MEKA: A Multi-label Extension to WEKA*, [Online].
- Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas (2010), "Mining Multi-label Data," *Data Mining and Knowledge Discovery Handbook*, hal. 1-20.
- Salvador García, Julian Luengo, and Francisco Herrera (2015), *Data Preprocessing in Data Mining*, Springer, Granada, Spain, Jaén, Spain,
- Serba linux, diakses 23 April 2018, <http://opensourceall.blogspot.co.id/2015/10/aplikasi-normalisasi-data.html>, Faisal Blach (2015), *Normalisasi Data pada Data Mining*, [Online].

Golden Helix, Inc, diakses 22 April 2018,  
<http://blog.goldenhelix.com/goldenadmin/cross-validation-for-genomic-prediction-in-svs/>, Golden Helix (2017), *Cross-Validation for Genomic Prediction in SVS*, [Online].

Ian H. Witten, and Eibe Frank (2005), *Data Mining : Practical Machine Learning Tools And Techniques*, 2nd edition, Morgan Kaufmann, San Francisco, USA.

P. Ravisankar, V. Ravi G. Raghava Rao, and I. Bose (2011), “Detection of financial statement fraud and feature selection using data mining techniques”, *Decision Support Systems*, vol. 50, hal. 491-500

## LAMPIRAN

Lampiran 1. Contoh Daftar Akun Pada Neraca Laporan Keuangan Pemerintah Daerah

No.	Daftar Akun Neraca
1	ASET
1.1	ASET LANCAR
1.1.1	Kas dan Setara Kas
1.1.1.1	Kas di Kas Daerah
1.1.1.2	Kas di Bendahara Pengeluaran
1.1.1.3	Kas di Bendahara Penerimaan
1.1.1.4	Kas di BLUD
1.1.1.5	Kas Dana JKN
1.1.1.6	Kas Dan BOS
1.1.1.7	Kas Lainnya
1.1.1.8	Kas yang Belum Dipertanggungjawabkan
1.1.1.9	Kas yang dibatasi Penggunaannya
1.1.2	Investasi Jangka Pendek
1.1.3	Piutang
1.1.3.1	Piutang Pajak
1.1.3.2	Piutang Retribusi
1.1.3.3	Piutang Sewa
1.1.3.4	Piutang Dana Bagi Hasil
1.1.3.5	Piutang Pendapatan Bagi Hasil dari Pusat
1.1.3.6	Piutang Pendapatan Bagi Hasil dari Provinsi
1.1.3.7	Piutang DAU
1.1.3.8	Piutang DAK
1.1.3.9	Penyisihan Piutang Tak Tertagih
1.1.3.10	Bagian Lancar Pinjaman kepada Perusahaan Negara
1.1.3.11	Bagian Lancar Pinjaman kepada Perusahaan Daerah
1.1.3.12	Bagian Lancar Pinjaman kepada Pemerintah Pusat
1.1.3.13	Bagian Lancar Pinjaman kepada Pemerintah Daerah Lainnya
1.1.3.14	Bagian Lancar Tagihan Penjualan Angsuran
1.1.3.15	Bagian Lancar Tuntutan Perbendaharaan
1.1.3.16	Bagian lancar Tuntutan Ganti Rugi
1.1.3.17	Piutang Lainnya
1.1.3.18	Piutang Lain-lain
1.1.3.19	Beban dibayar dimuka
1.1.4	Persediaan
	Jumlah Aset Lancar
1.2	INVESTASI JANGKA PANJANG
1.2.1	Investasi Nonpermanen
1.2.1.1	Pinjaman Kepada Perusahaan Negara
1.2.1.2	Pinjaman Kepada Perusahaan Daerah
1.2.1.3	Pinjaman Kepada Pemerintah Daerah Lainnya
1.2.1.4	Investasi dalam Surat Utang Negara
1.2.1.5	Investasi dalam Proyek Pembangunan
1.2.1.6	Investasi Nonpermanen Lainnya

No.	Daftar Akun Neraca
	Jumlah Investasi Nonpermanen
1.2.2	Investasi Permanen
1.2.2.1	Penyertaan Modal Pemerintah Daerah
1.2.2.2	Investasi Permanen Lainnya
	Jumlah Investasi Permanen
	Jumlah Investasi Jangka Panjang
1.3	ASET TETAP
1.3.1	Tanah
1.3.2	Peralatan dan Mesin
1.3.3	Gedung dan Bangunan
1.3.4	Jalan, Irigasi, dan Jaringan
1.3.5	Aset Tetap Lainnya
1.3.6	Konstruksi dalam Pengerjaan
1.3.7	Akumulasi Penyusutan
	Jumlah Aset Tetap
1.4	DANA CADANGAN
1.4.1	Dana Cadangan
	Jumlah Dana Cadangan
1.5	ASET LAINNYA
1.5.1	Tagihan Piutang Penjualan Angsuran
1.5.2	Tuntutan Perbendaharaan
1.5.3	Tuntutan Ganti Kerugian Daerah
1.5.4	Kemitraan dengan Pihak Ketiga
1.5.5	Aset Tak Berwujud
1.5.6	Aset Lain-Lain
1.5.7	Akumulasi Amortisasi ATB
	Kas yg dibatasi penggunaannya
1.5.8	Tagihan Jangka Panjang
	Jumlah Aset Lainnya
	<b>JUMLAH ASET</b>
2	KEWAJIBAN
2.1	KEWAJIBAN JANGKA PENDEK
2.1.1	Utang Perhitungan Pihak Ketiga (PFK)
2.1.2	Utang Belanja Pada Pihak Ketiga (PPK)
2.1.3	Utang Bunga Jangka Pendek
2.1.4	Utang Bunga Luar Negeri Jangka Pendek
2.1.5	Bagian Lancar Utang Dalam Negeri - Pemerintah Pusat
2.1.6	Bagian Lancar Utang Dalam Negeri- Pemerintah Daerah Lainnya
2.1.7	Bagian Lancar Utang Dalam Negeri - Lembaga Keuangan Bank
2.1.8	Bagian Lancar Utang Dalam Negeri - Lembaga Keuangan bukan Bank
2.1.9	Bagian Lancar Utang Dalam Negeri - Obligasi
2.1.10	Bagian Lancar Utang Jangka Panjang Lainnya
2.1.11	Utang Jangka Pendek Lainnya
2.1.12	Pendapatan Diterima Dimuka
2.1.13	Kewajiban/Hutang luar negeri jangka pendek
2.1.14	Utang Beban
	Jumlah Kewajiban Jangka Pendek
2.2	KEWAJIBAN JANGKA PANJANG

No.	Daftar Akun Neraca
2.2.1	Utang Dalam Negeri - Pemerintah Pusat
2.2.2	Utang Dalam Negeri - Pemerintah Daerah Lainnya
2.2.3	Utang Dalam Negeri - Lembaga Keuangan Bank
2.2.4	Utang Dalam Negeri - Lembaga Keuangan bukan Bank
2.2.5	Utang Dalam Negeri - Obligasi
2.2.6	Utang Dalam Negeri
2.2.7	Utang Luar Negeri /Kewajiban luar negeri jangka panjang
2.2.8	Utang Bunga Luar Negeri Jangka Panjang
2.2.9	Utang Bunga Jangka Panjang
2.2.10	Utang Jangka Panjang Lainnya
	Jumlah Kewajiban Jangka Panjang
	JUMLAH KEWAJIBAN
3	EKUITAS DANA
3.1	EKUITAS DANA LANCAR
3.1.1	Sisa Lebih Pembiayaan Anggaran (SiLPA)
3.1.2	Pendapatan yang Ditangguhkan
3.1.3	Cadangan Piutang
3.1.4	Cadangan Persediaan
3.1.5	Dana yang Harus Disediakan untuk Pembayaran Utang Jangka Pendek
	Jumlah Ekuitas Dana Lancar
3.2	EKUITAS DANA INVESTASI
3.2.1	Diinvestasikan dalam Investasi Jangka Panjang
3.2.2	Diinvestasikan dalam Aset Tetap
3.2.3	Diinvestasikan dalam Aset Lainnya
3.2.4	Dana yang Harus Disediakan untuk Pembayaran Utang Jangka Panjang
	Jumlah Ekuitas Dana Investasi
3.3	EKUITAS DANA CADANGAN
3.3.1	Diinvestasikan dalam Dana Cadangan
	Jumlah Ekuitas Dana Cadangan
	JUMLAH EKUITAS DANA
	JUMLAH KEWAJIBAN DAN EKUITAS DANA

*Halaman ini sengaja dikosongkan*



Lampiran 2. Contoh Daftar Akun Pada Laporan Realisasi Anggaran (LRA)  
Laporan Keuangan Pemerintah Daerah

No.	Daftar Akun LRA - Laporan Realisasi Anggaran
4	PENDAPATAN
4.1	PENDAPATAN ASLI DAERAH
4.1.1	Pendapatan Pajak Daerah
4.1.2	Pendapatan Retribusi Daerah
4.1.3	Pendapatan Hasil Pengelolaan Kekayaan Daerah yang Dipisahkan
4.1.4	Lain-lain PAD yang sah
	Jumlah Pendapatan Asli Daerah
4.2	PENDAPATAN TRANSFER
4.2.1	TRANSFER PEMERINTAH PUSAT - DANA PERIMBANGAN
4.2.1.1	Dana Bagi Hasil Pajak
4.2.1.2	Dana Bagi Hasil Bukan Pajak (Sumber Daya Alam)
4.2.1.3	Dana Alokasi Umum
4.2.1.4	Dana Alokasi Khusus
	Jumlah Pendapatan Transfer Dana Perimbangan
4.2.2	TRANSFER PEMERINTAH PUSAT - LAINNYA
4.2.2.1	Dana Otonomi Khusus
4.2.2.2	Dana Penyesuaian
	Jumlah Pendapatan Transfer Lainnya
4.2.3	BANTUAN KEUANGAN
4.2.3.1	Bantuan Keuangan dari Pemerintah Daerah Kabupaten
4.2.3.2	Bantuan Keuangan dari Pemerintah Daerah Provinsi
4.2.3.3	Bantuan Keuangan dari Pemerintah Daerah Kota
	Jumlah Bantuan Keuangan
	Total Pendapatan Transfer
4.3	LAIN-LAIN PENDAPATAN YANG SAH
4.3.1	Pendapatan Hibah
4.3.2	Dana Bagi Hasil Pajak dari Provinsi /Transfer dari Provinsi
4.3.3	Pendapatan Dana Darurat
4.3.4	Pendapatan Lainnya
	Jumlah Pendapatan Lain-lain yang Sah
	JUMLAH PENDAPATAN
5	BELANJA
5.1	BELANJA OPERASI
5.1.1	Belanja Pegawai
5.1.2	Belanja Barang (Barang dan Jasa)
5.1.3	Belanja Bunga
5.1.4	Belanja Subsidi
5.1.5	Belanja Hibah
5.1.6	Belanja Bantuan Sosial
5.1.7	Belanja Bantuan Keuangan
	Jumlah Belanja Operasi
5.2	BELANJA MODAL
5.2.1	Belanja Tanah
5.2.2	Belanja Peralatan dan Mesin
5.2.3	Belanja Gedung dan Bangunan
5.2.4	Belanja Jalan, Irigasi dan Jaringan

No.	Daftar Akun LRA - Laporan Realisasi Anggaran
5.2.5	Belanja Aset Tetap Lainnya
5.2.6	Belanja Aset Lainnya
	Jumlah Belanja Modal
5.3	BELANJA TAK TERDUGA
5.3.1	Belanja Tak Terduga
	Jumlah Belanja Tak Terduga
	JUMLAH BELANJA
6	TRANSFER
6.1	TRANSFER/BAGI HASIL PENDAPATAN KE KABUPATEN/KOTA
6.1.1	Bagi Hasil Pajak ke Kabupaten/Kota
6.1.2	Bagi Hasil Retribusi ke Kabupaten/Kota
6.1.3	Bagi Hasil Pendapatan Lainnya ke Kabupaten/Kota
	Jumlah Transfer Bagi Hasil Pendapatan ke Kab./Kota
	TRANSFER BANTUAN KEUANGAN
	Transfer Bantuan Keuangan ke Pemerintah Daerah Lainnya
	Transfer Bantuan Keuangan kepada Pemerintah Provinsi Kalimantan Tengah
	Transfer Bantuan Keuangan Lainnya
	Transfer Bantuan Keuangan ke Pemerintahan Desa
	Jumlah Transfer Bantuan Keuangan
	JUMLAH TRANSFER
	JUMLAH BELANJA DAN TRANSFER
7	SURPLUS/DEFISIT
8	PEMBIAYAAN
8.1	PENERIMAAN PEMBIAYAAN
8.1.1	Penggunaan SILPA
8.1.2	Pencairan Dana Cadangan
8.1.3	Hasil Penjualan Kekayaan Daerah yang Dipisahkan
8.1.4	Pinjaman Dalam Negeri - Pemerintah Pusat
8.1.5	Pinjaman Dalam Negeri - Pemerintah Daerah Lainnya
8.1.6	Pinjaman Dalam Negeri - Lembaga Keuangan Bank
8.1.7	Pinjaman Dalam Negeri - Lembaga Keuangan Bukan Bank
8.1.8	Pinjaman Dalam Negeri - Obligasi
8.1.9	Pinjaman Dalam Negeri - Lainnya
8.1.10	Pinjaman Dalam Negeri
8.1.11	Penerimaan Kembali Pinjaman kepada Perusahaan Negara
8.1.12	Penerimaan Kembali Pinjaman kepada Perusahaan Daerah
8.1.13	Penerimaan Kembali Pinjaman kepada Pemerintah Daerah Lainnya
	Jumlah Penerimaan Pembiayaan
8.2	PENGELUARAN PEMBIAYAAN
8.2.1	Pembentukan Dana Cadangan
8.2.2	Penyertaan Modal (Investasi) Pemerintah Daerah
8.2.3	Pembayaran Pokok Hutang/Kewajiban
8.2.4	Pembayaran Pokok Pinjaman Dalam Negeri - Pemerintah Pusat
8.2.5	Pembayaran Pokok Pinjaman Dalam Negeri - Pemerintah Daerah Lainnya
8.2.6	Pembayaran Pokok Pinjaman Dalam Negeri - Lembaga Keuangan Bank
8.2.7	Pembayaran Pokok Pinjaman Dalam Negeri - Lembaga Keuangan Bukan Bank
8.2.8	Pembayaran Pokok Pinjaman Dalam Negeri - Obligas
8.2.9	Pembayaran Pokok Pinjaman Dalam Negeri - Lainnya

No.	Daftar Akun LRA - Laporan Realisasi Anggaran
8.2.10	Pembayaran Pokok Pinjaman Dalam Negeri
8.2.11	Pemberian Pinjaman kepada Perusahaan Negara
8.2.12	Pemberian Pinjaman kepada Perusahaan Daerah
8.2.13	Pemberian Pinjaman kepada Pemerintah Daerah Lainnya
	Jumlah Pengeluaran Pembiayaan
8.3	PEMBIAYAAN NETO
9	Sisa Lebih Pembiayaan Anggaran (SiLPA)

*Halaman ini sengaja dikosongkan*

Lampiran 3. Contoh Daftar Akun Pada Laporan Arus Kas (LAK) Laporan Keuangan Pemerintah Daerah

No.	Daftar Akun LAK - Laporan Arus Kas
10	Arus Kas dari Aktivitas Operasi
10.1	Arus Masuk Kas dari Aktivitas Operasi
10.1.1	Pendapatan Pajak Daerah
10.1.2	Pendapatan Retribusi Daerah
10.1.3	Pendapatan Hasil Pengelolaan Kekayaan Daerah yang Dipisahkan
10.1.4	Lain-lain PAD yang sah
10.1.5	Dana Bagi Hasil Pajak
10.1.6	Dana Bagi Hasil Sumber Daya Alam
10.1.7	Dana Alokasi Umum
10.1.8	Dana Alokasi Khusus
10.1.9	Dana Otonomi Khusus
10.1.10	Dana Penyesuaian
10.1.11	Penerimaan Transfer Pemerintah Pusat
10.1.12	Penerimaan Transfer Pemerintah Pusat – Lainnya
10.1.13	Penerimaan Bantuan Keuangan
10.1.14	Pendapatan Hibah
10.1.15	Pendapatan Dana Darurat
10.1.16	Dana Bagi Hasil Pajak dari Provinsi
10.1.17	Pendapatan Lainnya
	Jumlah Arus Masuk Kas dari Aktivitas Operasi
10.2	Arus Keluar Kas dari Aktivitas Operasi
10.2.1	Belanja Pegawai
10.2.2	Belanja Barang
10.2.3	Bunga
10.2.4	Subsidi
10.2.5	Hibah
10.2.6	Bantuan Sosial dan Politik
10.2.7	Bantuan Keuangan
10.2.8	Belanja Tak Terduga
10.2.9	Bagi Hasil Pajak ke Kabupaten/Kota
10.2.10	Bagi Hasil Retribusi ke Kabupaten/Kota
10.2.11	Bagi Hasil Pendapatan Lainnya ke Kabupaten/Kota
10.2.12	Pembayaran Bantuan Keuangan ke Pemerintah Daerah Lainnya
10.2.13	Pembayaran Bantuan Keuangan Lainnya
10.2.14	Pembayaran Bantuan Keuangan ke Pemerintahan Provinsi Kalimantan Tengah
10.2.15	Pembayaran Bantuan Keuangan ke Pemerintahan Desa
10.2.16	Transfer /Bagi Hasil ke Kabupaten/Kota
	Jumlah Arus Keluar Kas dari Aktivitas Operasi
	Arus Kas Bersih dari Aktivitas Operasi
11	Arus Kas dari Aktivitas Investasi Aset Non keuangan
11.1	Arus Masuk Kas dari Aktivitas Investasi Aset Non keuangan
11.1.1	Pendapatan Penjualan atas Tanah
11.1.2	Pendapatan Penjualan atas Peralatan dan Mesin
11.1.3	Pendapatan Penjualan atas Gedung dan Bangunan
11.1.4	Pendapatan Penjualan atas Jalan, Irigasi dan Jaringan
11.1.5	Pendapatan dari Penjualan Aset Tetap Lainnya

No.	Daftar Akun LAK - Laporan Arus Kas
11.1.6	Pendapatan dari Penjualan Aset Lainnya
11.1.7	Pencairan Dana Cadangan
11.1.8	Penjualan Rumah Jabatan/Rumah Dinas
11.1.9	Penjualan Kendaraan Dinas roda dua
11.1.10	Penjualan Kendaraan Dinas roda empat
	Jumlah Arus Masuk Kas dari Aktivitas Investasi Aset Non keuangan
11.2	Arus Keluar Kas dari Aktivitas Investasi Aset Non keuangan
11.2.1	Belanja Tanah
11.2.2	Belanja Peralatan dan Mesin
11.2.3	Belanja Gedung dan Bangunan
11.2.4	Belanja Jalan, Irigasi dan Jaringan
11.2.5	Belanja Aset Tetap Lainnya
11.2.6	Penyertaan Modal/Investasi Pemerintah Daerah
11.2.7	Belanja Aset Lainnya
	Jumlah Arus Keluar Kas dari Aktivitas Investasi Aset Non keuangan
	Arus Kas Bersih dari Aktivitas Investasi Aset Non keuangan
12	Arus Kas dari Aktivitas Pembiayaan
12.1	Arus Masuk Kas dari Aktivitas Pembiayaan
12.1.1	Pencairan Dana Cadangan
12.1.2	Hasil Penjualan Kekayaan Daerah yang Dipisahkan
12.1.3	Pinjaman Dalam Negeri - Pemerintah Pusat
12.1.4	Pinjaman Dalam Negeri - Pemerintah Daerah Lainnya
12.1.5	Pinjaman Dalam Negeri - Lembaga Keuangan Bank
12.1.6	Pinjaman Dalam Negeri - Lembaga Keuangan Bukan Bank
12.1.7	Pinjaman Dalam Negeri - Obligasi
12.1.8	Pinjaman Dalam Negeri - Lainnya
12.1.9	Penerimaan Kembali Pinjaman kepada Perusahaan Negara
12.1.10	Penerimaan Kembali Pinjaman kepada Perusahaan Daerah
12.1.11	Penerimaan Kembali Pinjaman kepada Pemerintah Daerah Lainnya
12.1.12	Saldo Awal JKN
	Sisa Perhitungan Tahun Lalu
12.1.13	Penerimaan UP/UUDP Tahun Lalu
	Jumlah Arus Masuk Kas dari Aktivitas Pembiayaan
12.2	Arus Keluar Kas dari Aktivitas Pembiayaan
12.2.1	Pembentukan Dana Cadangan
12.2.2	Penyertaan Modal Pemerintah Daerah
12.2.3	Pembayaran Pokok Utang
12.2.4	Pembayaran Pokok Pinjaman Dalam Negeri - Pemerintah Pusat
12.2.5	Pembayaran Pokok Pinjaman Dalam Negeri - Pemerintah Daerah Lainnya
12.2.6	Pembayaran Pokok Pinjaman Dalam Negeri - Lembaga Keuangan Bank
12.2.7	Pembayaran Pokok Pinjaman Dalam Negeri - Lembaga Keuangan Bukan Bank
12.2.8	Pembayaran Pokok Pinjaman Dalam Negeri - Obligasi
12.2.9	Pembayaran Pokok Pinjaman Dalam Negeri - Lainnya
12.2.10	Pemberian Pinjaman kepada Perusahaan Negara
12.2.11	Pemberian Pinjaman kepada Perusahaan Daerah
12.2.12	Pemberian Pinjaman kepada Pemerintah Daerah Lainnya
	Jumlah Arus Keluar Kas dari Aktivitas Pembiayaan
	Arus Kas Bersih dari Aktivitas Pembiayaan

No.	Daftar Akun LAK - Laporan Arus Kas
13	Arus Kas dari Aktivitas Non Anggaran
13.1	Arus Masuk Kas dari Aktivitas Non Anggaran
13.1.1	Penerimaan Perhitungan Pihak Ketiga (PFK)
13.1.2	Pendapatan Yang Belum Masuk Bank
13.1.3	Penerimaan UUDP Tahun Lalu
13.1.4	Setoran Kas di Bendahara Penerimaan
	Sisa Dana JKN TA Sebelumnya
	Sisa Dana BOS TA Sebelumnya
	Penerimaan Urusan Keuangan dan Perhitungan (UKP)
13.1.5	Kiriman Uang Masuk /Kas di Bendahara Pengeluaran Tahun Sebelumnya
	Jumlah Arus Masuk Kas dari Aktivitas Non Anggaran
13.2	Arus Keluar Kas dari Aktivitas Non Anggaran
13.2.1	Pengeluaran Perhitungan Pihak Ketiga (PFK)
13.2.2	Pengeluaran BG Yang Belum Dicairkan di Bank
13.2.3	Pengeluaran Kekurangan Bayar SP2D pada Pihak Ketiga
13.2.4	Sisa Dana JKN
13.2.5	Sisa Dana BOS
	Sisa Dana Bansos Pendidikan
	Pengeluaran Urusan Keuangan dan Perhitungan (UKP)
13.2.6	Kiriman Uang Keluar /Sisa Kas di Bendahara Penegeluaran Tahun Berjalan
	Jumlah Arus Keluar Kas dari Aktivitas Non Anggaran
	Arus Kas Bersih dari Aktivitas Nonanggaran
14	Kenaikan/Penurunan Kas
15	Saldo Awal Kas di BUD /Kas Daerah
16	Saldo Akhir Kas (Sebelum Koreksi) - di BUD/Kas Daerah
17	Koreksi atas nilai Kas Bendahara Pengeluaran
18	Saldo Akhir Kas di BUD/Kas Daerah
19	Saldo Akhir Kas di Bendahara Pengeluaran
20	Saldo Akhir Kas di Bendahara Penerimaan
21	Saldo Akhir Kas di BLUD
22	Saldo Akhir Kas Dana JKN
23	Saldo Akhir Kas Dana BOS
24	Saldo Akhir Kas Lainnya
25	Utang PFK
26	Saldo Akhir Kas yang Dibatasi Penggunaannya
26	Saldo AKHIR KAS

*Halaman ini sengaja dikosongkan*



Lampiran 4. Hasil *Hamming Loss* pada setiap dataset

Hamming Loss			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	1	0.270	0.240	0.242	0.503	0.286	0.248	0.243	0.485	0.287	0.286	0.278	0.321
		2	0.270	0.239	0.232	0.494	0.286	0.254	0.233	0.473	0.295	0.294	0.270	0.316
		3	0.267	0.231	0.234	0.489	0.297	0.243	0.235	0.477	0.293	0.282	0.279	0.319
		4	0.257	0.233	0.226	0.499	0.274	0.250	0.248	0.489	0.294	0.278	0.273	0.317
		5	0.235	0.231	0.236	0.484	0.271	0.247	0.240	0.475	0.298	0.287	0.264	0.319
		6	0.265	0.230	0.226	0.469	0.314	0.240	0.229	0.459	0.311	0.284	0.265	0.321
		7	0.258	0.230	0.231	0.495	0.280	0.253	0.238	0.473	0.296	0.289	0.266	0.322
		8	0.250	0.232	0.240	0.506	0.302	0.246	0.241	0.486	0.309	0.286	0.275	0.309
		9	0.259	0.236	0.231	0.482	0.290	0.245	0.231	0.460	0.310	0.286	0.277	0.316
		10	0.260	0.234	0.231	0.490	0.290	0.252	0.231	0.474	0.304	0.281	0.258	0.309
2	F1_N2	1	0.266	0.238	0.246	0.501	0.281	0.247	0.246	0.486	0.310	0.284	0.275	0.314
		2	0.270	0.235	0.244	0.492	0.290	0.253	0.236	0.473	0.272	0.294	0.275	0.317
		3	0.262	0.236	0.238	0.494	0.286	0.246	0.238	0.482	0.276	0.278	0.269	0.316
		4	0.263	0.237	0.231	0.501	0.285	0.249	0.233	0.486	0.298	0.282	0.270	0.316
		5	0.250	0.233	0.236	0.481	0.301	0.246	0.233	0.473	0.299	0.283	0.266	0.315
		6	0.246	0.233	0.231	0.472	0.279	0.239	0.234	0.453	0.275	0.289	0.275	0.322
		7	0.253	0.232	0.235	0.495	0.286	0.254	0.237	0.476	0.280	0.290	0.279	0.324
		8	0.256	0.235	0.229	0.503	0.287	0.245	0.235	0.488	0.286	0.286	0.281	0.310
		9	0.270	0.234	0.236	0.478	0.280	0.246	0.231	0.463	0.290	0.289	0.281	0.321
		10	0.264	0.234	0.234	0.488	0.287	0.251	0.238	0.470	0.292	0.284	0.262	0.311
3	F1_N3	1	0.268	0.238	0.234	0.496	0.281	0.245	0.238	0.483	0.301	0.285	0.268	0.322
		2	0.268	0.235	0.237	0.494	0.293	0.256	0.241	0.473	0.285	0.300	0.276	0.320
		3	0.257	0.234	0.242	0.491	0.285	0.243	0.241	0.482	0.284	0.282	0.269	0.324
		4	0.250	0.231	0.237	0.496	0.281	0.249	0.238	0.484	0.315	0.282	0.263	0.314
		5	0.243	0.232	0.235	0.484	0.303	0.247	0.232	0.470	0.289	0.288	0.267	0.311
		6	0.243	0.231	0.226	0.470	0.274	0.241	0.236	0.458	0.320	0.282	0.275	0.321
		7	0.251	0.232	0.240	0.490	0.292	0.253	0.235	0.472	0.288	0.290	0.266	0.319
		8	0.253	0.233	0.237	0.506	0.285	0.245	0.231	0.491	0.288	0.289	0.282	0.315
		9	0.267	0.235	0.231	0.478	0.273	0.248	0.229	0.465	0.288	0.288	0.270	0.318
		10	0.255	0.234	0.232	0.490	0.291	0.250	0.232	0.476	0.293	0.277	0.266	0.311
4	F1_N4	1	0.259	0.220	0.242	0.458	0.280	0.228	0.235	0.430	0.299	0.272	0.283	0.308
		2	0.260	0.222	0.234	0.447	0.291	0.234	0.240	0.430	0.285	0.268	0.280	0.301
		3	0.261	0.214	0.241	0.438	0.296	0.231	0.239	0.429	0.287	0.267	0.264	0.297
		4	0.256	0.222	0.233	0.438	0.278	0.230	0.237	0.432	0.313	0.264	0.264	0.303
		5	0.250	0.225	0.235	0.448	0.310	0.230	0.247	0.433	0.299	0.250	0.283	0.296
		6	0.243	0.220	0.230	0.428	0.267	0.233	0.229	0.421	0.298	0.270	0.280	0.307
		7	0.250	0.210	0.233	0.436	0.297	0.230	0.245	0.427	0.274	0.255	0.270	0.317
		8	0.250	0.218	0.241	0.449	0.285	0.225	0.233	0.431	0.292	0.259	0.276	0.305
		9	0.265	0.222	0.232	0.424	0.282	0.229	0.234	0.413	0.286	0.265	0.272	0.303
		10	0.260	0.220	0.230	0.445	0.279	0.226	0.225	0.430	0.290	0.253	0.272	0.313
5	F1_N5	1	0.262	0.242	0.242	0.417	0.280	0.250	0.237	0.403	0.331	0.274	0.287	0.312
		2	0.270	0.238	0.230	0.400	0.290	0.247	0.238	0.399	0.299	0.298	0.264	0.299
		3	0.270	0.238	0.226	0.408	0.310	0.241	0.230	0.395	0.308	0.287	0.270	0.300
		4	0.256	0.245	0.227	0.415	0.285	0.248	0.237	0.399	0.297	0.278	0.261	0.314
		5	0.249	0.235	0.235	0.400	0.267	0.249	0.240	0.390	0.311	0.275	0.288	0.295
		6	0.257	0.240	0.227	0.393	0.293	0.239	0.234	0.390	0.323	0.280	0.276	0.304
		7	0.263	0.232	0.233	0.399	0.276	0.251	0.243	0.390	0.316	0.272	0.278	0.306
		8	0.257	0.242	0.231	0.414	0.287	0.241	0.234	0.408	0.310	0.282	0.278	0.312
		9	0.268	0.241	0.232	0.394	0.295	0.250	0.234	0.391	0.290	0.284	0.276	0.298
		10	0.257	0.242	0.230	0.405	0.295	0.243	0.236	0.403	0.300	0.275	0.255	0.308
6	F2_N1	1	0.219	0.230	0.209	0.304	0.215	0.272	0.197	0.340	0.274	0.310	0.226	0.258
		2	0.216	0.227	0.193	0.306	0.227	0.266	0.195	0.330	0.253	0.302	0.210	0.245
		3	0.210	0.233	0.195	0.322	0.226	0.273	0.199	0.354	0.251	0.314	0.200	0.260

Hamming Loss			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
		4	0.224	0.232	0.199	0.309	0.223	0.269	0.199	0.331	0.240	0.313	0.230	0.268
		5	0.208	0.231	0.197	0.321	0.219	0.271	0.200	0.351	0.229	0.310	0.223	0.272
		6	0.222	0.232	0.193	0.310	0.217	0.266	0.203	0.326	0.255	0.316	0.211	0.268
		7	0.211	0.229	0.194	0.323	0.214	0.274	0.192	0.331	0.233	0.316	0.211	0.268
		8	0.214	0.231	0.200	0.310	0.219	0.276	0.198	0.346	0.238	0.313	0.233	0.268
		9	0.214	0.232	0.184	0.313	0.197	0.272	0.182	0.351	0.266	0.320	0.214	0.267
		10	0.213	0.230	0.192	0.309	0.210	0.271	0.196	0.344	0.253	0.307	0.223	0.279
7	F2_N2	1	0.219	0.233	0.206	0.306	0.215	0.271	0.205	0.344	0.271	0.309	0.230	0.257
		2	0.216	0.227	0.200	0.305	0.227	0.266	0.202	0.330	0.243	0.304	0.210	0.250
		3	0.213	0.235	0.199	0.321	0.226	0.275	0.190	0.354	0.251	0.316	0.207	0.266
		4	0.225	0.231	0.192	0.310	0.223	0.268	0.191	0.332	0.238	0.313	0.228	0.267
		5	0.207	0.232	0.195	0.320	0.218	0.270	0.198	0.350	0.246	0.310	0.219	0.270
		6	0.221	0.231	0.201	0.311	0.219	0.265	0.193	0.327	0.250	0.316	0.226	0.264
		7	0.211	0.229	0.201	0.324	0.214	0.274	0.203	0.333	0.230	0.318	0.223	0.265
		8	0.216	0.234	0.193	0.311	0.219	0.278	0.195	0.346	0.233	0.313	0.235	0.270
		9	0.215	0.231	0.192	0.315	0.198	0.271	0.190	0.351	0.266	0.320	0.214	0.263
		10	0.213	0.230	0.190	0.309	0.210	0.271	0.204	0.344	0.251	0.307	0.227	0.283
8	F2_N3	1	0.219	0.230	0.204	0.303	0.215	0.271	0.201	0.340	0.271	0.309	0.234	0.256
		2	0.216	0.227	0.192	0.303	0.227	0.266	0.196	0.329	0.243	0.302	0.220	0.243
		3	0.213	0.234	0.190	0.324	0.226	0.273	0.194	0.356	0.253	0.316	0.218	0.263
		4	0.225	0.232	0.190	0.309	0.223	0.267	0.198	0.328	0.238	0.313	0.224	0.264
		5	0.207	0.231	0.193	0.319	0.219	0.270	0.200	0.354	0.249	0.312	0.226	0.272
		6	0.221	0.231	0.192	0.309	0.219	0.266	0.191	0.329	0.250	0.316	0.214	0.264
		7	0.211	0.229	0.190	0.324	0.214	0.274	0.192	0.328	0.230	0.316	0.215	0.261
		8	0.216	0.234	0.211	0.310	0.219	0.276	0.201	0.347	0.234	0.317	0.233	0.266
		9	0.215	0.231	0.188	0.314	0.197	0.272	0.185	0.350	0.266	0.316	0.218	0.255
		10	0.213	0.230	0.188	0.310	0.210	0.271	0.199	0.340	0.251	0.312	0.227	0.269
9	F2_N4	1	0.219	0.204	0.209	0.249	0.215	0.232	0.206	0.269	0.271	0.245	0.222	0.240
		2	0.216	0.203	0.198	0.251	0.227	0.230	0.200	0.268	0.243	0.234	0.235	0.232
		3	0.213	0.204	0.196	0.259	0.226	0.226	0.201	0.279	0.253	0.240	0.214	0.247
		4	0.225	0.204	0.193	0.250	0.223	0.230	0.197	0.270	0.238	0.250	0.223	0.238
		5	0.207	0.216	0.205	0.261	0.219	0.235	0.195	0.289	0.249	0.248	0.223	0.240
		6	0.221	0.201	0.195	0.254	0.219	0.224	0.197	0.269	0.250	0.235	0.215	0.242
		7	0.210	0.195	0.204	0.249	0.214	0.228	0.190	0.272	0.230	0.235	0.211	0.241
		8	0.216	0.204	0.208	0.253	0.219	0.234	0.202	0.271	0.234	0.251	0.240	0.248
		9	0.215	0.208	0.186	0.267	0.197	0.233	0.196	0.279	0.266	0.245	0.220	0.246
		10	0.213	0.202	0.195	0.260	0.210	0.237	0.191	0.278	0.251	0.250	0.237	0.248
10	F2_N5	1	0.218	0.232	0.202	0.264	0.214	0.270	0.204	0.296	0.275	0.310	0.229	0.246
		2	0.216	0.221	0.200	0.266	0.235	0.266	0.206	0.302	0.254	0.310	0.222	0.238
		3	0.210	0.227	0.193	0.274	0.226	0.270	0.194	0.310	0.254	0.320	0.210	0.248
		4	0.224	0.221	0.193	0.264	0.221	0.265	0.193	0.297	0.246	0.315	0.207	0.252
		5	0.208	0.229	0.198	0.270	0.220	0.267	0.196	0.300	0.229	0.309	0.229	0.255
		6	0.220	0.229	0.192	0.263	0.217	0.265	0.193	0.284	0.253	0.314	0.212	0.258
		7	0.210	0.223	0.193	0.263	0.214	0.265	0.199	0.285	0.233	0.320	0.210	0.246
		8	0.214	0.230	0.198	0.271	0.219	0.273	0.195	0.300	0.241	0.313	0.230	0.269
		9	0.215	0.225	0.192	0.271	0.197	0.269	0.195	0.308	0.268	0.322	0.212	0.265
		10	0.213	0.228	0.193	0.272	0.209	0.270	0.193	0.300	0.255	0.314	0.230	0.261

Lampiran 5. Hasil *One-Error* pada setiap dataset

One-Error			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	1	0.560	0.467	0.360	0.693	0.487	0.433	0.447	0.653	0.593	0.547	0.540	0.633
		2	0.533	0.433	0.387	0.647	0.547	0.440	0.400	0.627	0.587	0.567	0.480	0.600
		3	0.547	0.433	0.407	0.660	0.567	0.413	0.400	0.627	0.607	0.553	0.500	0.620
		4	0.500	0.427	0.413	0.673	0.513	0.433	0.460	0.660	0.533	0.527	0.493	0.633
		5	0.433	0.413	0.347	0.640	0.467	0.433	0.447	0.633	0.580	0.553	0.453	0.600
		6	0.547	0.420	0.360	0.667	0.613	0.407	0.413	0.640	0.607	0.547	0.480	0.620
		7	0.513	0.407	0.420	0.653	0.533	0.440	0.433	0.620	0.640	0.553	0.493	0.620
		8	0.480	0.420	0.393	0.673	0.553	0.420	0.460	0.640	0.593	0.547	0.500	0.627
		9	0.473	0.420	0.333	0.660	0.600	0.420	0.393	0.620	0.613	0.553	0.500	0.620
		10	0.607	0.413	0.373	0.673	0.573	0.427	0.413	0.620	0.593	0.533	0.473	0.593
2	F1_N2	1	0.553	0.460	0.407	0.680	0.540	0.433	0.473	0.660	0.627	0.540	0.533	0.607
		2	0.520	0.413	0.380	0.640	0.540	0.440	0.467	0.627	0.527	0.573	0.520	0.600
		3	0.540	0.440	0.380	0.653	0.527	0.427	0.427	0.627	0.560	0.540	0.467	0.607
		4	0.553	0.427	0.373	0.680	0.573	0.433	0.427	0.660	0.620	0.540	0.493	0.613
		5	0.520	0.427	0.360	0.640	0.547	0.433	0.420	0.647	0.580	0.540	0.480	0.607
		6	0.520	0.440	0.347	0.660	0.520	0.400	0.420	0.640	0.513	0.567	0.513	0.640
		7	0.533	0.420	0.400	0.647	0.533	0.440	0.433	0.633	0.553	0.553	0.520	0.627
		8	0.460	0.433	0.347	0.653	0.487	0.420	0.407	0.647	0.567	0.547	0.493	0.620
		9	0.560	0.413	0.380	0.660	0.567	0.427	0.387	0.627	0.547	0.560	0.513	0.620
		10	0.560	0.413	0.367	0.680	0.553	0.433	0.433	0.620	0.573	0.547	0.493	0.593
3	F1_N3	1	0.573	0.453	0.393	0.693	0.527	0.427	0.433	0.660	0.613	0.547	0.507	0.620
		2	0.567	0.407	0.393	0.640	0.567	0.447	0.447	0.620	0.553	0.587	0.500	0.600
		3	0.547	0.440	0.347	0.660	0.547	0.420	0.433	0.633	0.587	0.553	0.493	0.627
		4	0.553	0.420	0.393	0.687	0.533	0.433	0.433	0.653	0.607	0.540	0.473	0.620
		5	0.520	0.420	0.380	0.653	0.560	0.433	0.393	0.640	0.560	0.553	0.480	0.580
		6	0.493	0.420	0.367	0.667	0.533	0.407	0.433	0.640	0.607	0.547	0.513	0.647
		7	0.540	0.427	0.407	0.647	0.560	0.440	0.407	0.627	0.587	0.560	0.480	0.600
		8	0.493	0.427	0.373	0.667	0.520	0.420	0.413	0.653	0.580	0.553	0.527	0.613
		9	0.527	0.420	0.373	0.667	0.580	0.433	0.407	0.627	0.553	0.560	0.493	0.627
		10	0.573	0.413	0.387	0.693	0.580	0.427	0.373	0.620	0.580	0.527	0.493	0.593
4	F1_N4	1	0.533	0.393	0.400	0.600	0.513	0.387	0.427	0.613	0.587	0.460	0.553	0.620
		2	0.540	0.393	0.387	0.573	0.553	0.380	0.433	0.580	0.520	0.460	0.513	0.573
		3	0.553	0.347	0.407	0.560	0.593	0.380	0.420	0.587	0.573	0.447	0.460	0.573
		4	0.520	0.367	0.373	0.607	0.560	0.380	0.453	0.587	0.607	0.447	0.467	0.600
		5	0.553	0.387	0.387	0.593	0.593	0.393	0.473	0.620	0.600	0.407	0.513	0.580
		6	0.527	0.367	0.373	0.600	0.487	0.373	0.400	0.580	0.600	0.467	0.500	0.587
		7	0.513	0.333	0.407	0.607	0.567	0.380	0.453	0.580	0.587	0.427	0.480	0.613
		8	0.473	0.353	0.353	0.607	0.513	0.373	0.413	0.613	0.573	0.440	0.500	0.620
		9	0.560	0.367	0.393	0.593	0.580	0.360	0.387	0.587	0.540	0.453	0.487	0.587
		10	0.553	0.387	0.373	0.587	0.560	0.367	0.400	0.593	0.580	0.420	0.507	0.587
5	F1_N5	1	0.533	0.453	0.387	0.573	0.527	0.453	0.420	0.653	0.627	0.507	0.533	0.607
		2	0.533	0.440	0.400	0.560	0.533	0.427	0.440	0.627	0.587	0.587	0.513	0.607
		3	0.580	0.433	0.347	0.547	0.573	0.420	0.387	0.633	0.613	0.560	0.467	0.573
		4	0.487	0.473	0.360	0.587	0.547	0.433	0.440	0.640	0.620	0.527	0.467	0.593
		5	0.473	0.440	0.373	0.553	0.487	0.433	0.427	0.607	0.613	0.513	0.507	0.567
		6	0.500	0.453	0.353	0.560	0.507	0.413	0.413	0.620	0.640	0.527	0.500	0.560
		7	0.533	0.413	0.400	0.567	0.513	0.440	0.460	0.607	0.633	0.513	0.500	0.587
		8	0.533	0.460	0.393	0.600	0.560	0.420	0.427	0.647	0.633	0.547	0.507	0.620
		9	0.513	0.453	0.353	0.547	0.553	0.440	0.413	0.607	0.607	0.540	0.513	0.553
		10	0.547	0.480	0.387	0.553	0.573	0.413	0.413	0.613	0.573	0.520	0.500	0.567
6	F2_N1	1	0.420	0.393	0.293	0.473	0.333	0.507	0.340	0.600	0.513	0.647	0.387	0.533
		2	0.353	0.387	0.267	0.507	0.347	0.493	0.320	0.607	0.440	0.620	0.367	0.513
		3	0.347	0.413	0.267	0.507	0.333	0.513	0.340	0.640	0.473	0.653	0.320	0.547

One-Error			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
		4	0.380	0.407	0.273	0.520	0.307	0.507	0.347	0.613	0.453	0.660	0.400	0.567
		5	0.360	0.393	0.267	0.520	0.360	0.507	0.320	0.633	0.447	0.647	0.387	0.573
		6	0.393	0.413	0.280	0.487	0.340	0.487	0.353	0.587	0.453	0.667	0.380	0.567
		7	0.367	0.373	0.273	0.500	0.313	0.507	0.313	0.607	0.413	0.653	0.353	0.573
		8	0.420	0.387	0.273	0.533	0.313	0.513	0.340	0.647	0.480	0.647	0.400	0.573
		9	0.353	0.400	0.253	0.473	0.287	0.527	0.267	0.620	0.500	0.687	0.367	0.547
		10	0.380	0.387	0.267	0.540	0.333	0.507	0.327	0.613	0.447	0.633	0.367	0.573
7	F2_N2	1	0.420	0.400	0.287	0.487	0.333	0.507	0.380	0.600	0.507	0.640	0.440	0.540
		2	0.353	0.387	0.273	0.507	0.347	0.493	0.353	0.607	0.427	0.627	0.380	0.533
		3	0.347	0.413	0.273	0.507	0.333	0.513	0.320	0.640	0.473	0.660	0.373	0.560
		4	0.393	0.407	0.287	0.520	0.307	0.507	0.327	0.607	0.460	0.660	0.407	0.553
		5	0.367	0.393	0.287	0.520	0.360	0.507	0.307	0.633	0.480	0.647	0.380	0.567
		6	0.393	0.407	0.293	0.493	0.340	0.487	0.313	0.587	0.440	0.667	0.373	0.560
		7	0.367	0.373	0.293	0.507	0.313	0.507	0.333	0.607	0.407	0.660	0.373	0.553
		8	0.427	0.400	0.273	0.527	0.313	0.520	0.333	0.647	0.493	0.647	0.427	0.567
		9	0.347	0.393	0.253	0.480	0.287	0.527	0.307	0.627	0.500	0.687	0.353	0.540
		10	0.380	0.387	0.260	0.533	0.333	0.507	0.347	0.613	0.440	0.633	0.387	0.573
8	F2_N3	1	0.420	0.393	0.320	0.500	0.333	0.507	0.340	0.600	0.507	0.640	0.407	0.540
		2	0.353	0.387	0.273	0.507	0.347	0.493	0.347	0.607	0.427	0.620	0.387	0.500
		3	0.347	0.420	0.267	0.507	0.333	0.513	0.280	0.647	0.473	0.660	0.353	0.567
		4	0.393	0.413	0.287	0.527	0.307	0.507	0.347	0.607	0.460	0.660	0.400	0.533
		5	0.367	0.393	0.253	0.527	0.353	0.507	0.327	0.647	0.487	0.653	0.387	0.587
		6	0.393	0.407	0.287	0.500	0.340	0.487	0.300	0.600	0.440	0.667	0.373	0.553
		7	0.367	0.373	0.260	0.500	0.307	0.507	0.313	0.607	0.407	0.653	0.367	0.547
		8	0.427	0.400	0.287	0.527	0.313	0.513	0.347	0.653	0.500	0.660	0.420	0.560
		9	0.347	0.393	0.247	0.480	0.287	0.527	0.287	0.633	0.500	0.673	0.360	0.507
		10	0.380	0.387	0.247	0.540	0.333	0.507	0.320	0.607	0.440	0.653	0.360	0.540
9	F2_N4	1	0.420	0.347	0.327	0.333	0.333	0.373	0.360	0.467	0.507	0.407	0.407	0.507
		2	0.353	0.360	0.260	0.347	0.347	0.353	0.340	0.447	0.427	0.380	0.413	0.460
		3	0.347	0.360	0.267	0.353	0.333	0.340	0.327	0.460	0.473	0.400	0.380	0.520
		4	0.393	0.360	0.280	0.347	0.307	0.373	0.333	0.480	0.460	0.433	0.400	0.473
		5	0.367	0.427	0.300	0.360	0.353	0.380	0.313	0.500	0.487	0.413	0.373	0.480
		6	0.393	0.347	0.287	0.360	0.340	0.340	0.333	0.447	0.440	0.387	0.347	0.507
		7	0.360	0.327	0.260	0.327	0.307	0.353	0.293	0.467	0.407	0.387	0.340	0.493
		8	0.427	0.367	0.280	0.360	0.313	0.367	0.360	0.453	0.500	0.433	0.440	0.513
		9	0.347	0.387	0.267	0.380	0.287	0.367	0.320	0.487	0.500	0.407	0.360	0.487
		10	0.380	0.360	0.260	0.373	0.333	0.367	0.300	0.473	0.440	0.420	0.400	0.473
10	F2_N5	1	0.420	0.393	0.300	0.393	0.333	0.500	0.347	0.520	0.513	0.653	0.380	0.527
		2	0.353	0.373	0.273	0.420	0.373	0.487	0.353	0.520	0.440	0.647	0.373	0.513
		3	0.347	0.380	0.260	0.400	0.333	0.493	0.327	0.507	0.480	0.673	0.333	0.507
		4	0.380	0.380	0.267	0.393	0.300	0.500	0.340	0.500	0.467	0.667	0.360	0.513
		5	0.367	0.400	0.267	0.407	0.367	0.493	0.313	0.513	0.447	0.640	0.393	0.540
		6	0.393	0.373	0.287	0.407	0.340	0.487	0.313	0.493	0.453	0.660	0.360	0.547
		7	0.367	0.360	0.267	0.360	0.313	0.480	0.313	0.473	0.407	0.667	0.367	0.500
		8	0.420	0.387	0.260	0.380	0.313	0.500	0.320	0.513	0.493	0.647	0.413	0.580
		9	0.347	0.367	0.247	0.433	0.287	0.513	0.320	0.513	0.513	0.693	0.347	0.527
		10	0.380	0.387	0.273	0.427	0.327	0.500	0.300	0.520	0.453	0.660	0.407	0.527

Lampiran 6. Hasil *Rank Loss* pada setiap dataset

Rank Loss			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	1	0.337	0.409	0.229	0.458	0.417	0.403	0.384	0.476	0.396	0.434	0.416	0.443
		2	0.341	0.400	0.216	0.441	0.399	0.409	0.360	0.461	0.417	0.454	0.388	0.436
		3	0.384	0.386	0.227	0.432	0.420	0.388	0.360	0.465	0.398	0.439	0.368	0.433
		4	0.342	0.404	0.211	0.451	0.363	0.409	0.389	0.485	0.389	0.430	0.396	0.443
		5	0.266	0.390	0.198	0.440	0.348	0.395	0.368	0.461	0.366	0.436	0.361	0.422
		6	0.359	0.370	0.210	0.432	0.462	0.380	0.347	0.445	0.434	0.438	0.380	0.423
		7	0.339	0.386	0.224	0.444	0.382	0.408	0.369	0.453	0.376	0.440	0.378	0.435
		8	0.337	0.379	0.232	0.452	0.401	0.389	0.374	0.471	0.417	0.436	0.402	0.413
		9	0.316	0.391	0.199	0.427	0.420	0.396	0.354	0.449	0.391	0.439	0.391	0.418
		10	0.338	0.391	0.212	0.447	0.419	0.408	0.356	0.474	0.396	0.429	0.352	0.422
2	F1_N2	1	0.326	0.409	0.235	0.457	0.430	0.400	0.389	0.476	0.425	0.430	0.400	0.420
		2	0.345	0.408	0.220	0.440	0.403	0.408	0.370	0.457	0.352	0.454	0.379	0.431
		3	0.334	0.400	0.225	0.433	0.406	0.399	0.372	0.465	0.365	0.430	0.365	0.428
		4	0.343	0.407	0.214	0.457	0.408	0.411	0.349	0.484	0.401	0.437	0.380	0.427
		5	0.314	0.390	0.214	0.439	0.421	0.395	0.354	0.464	0.401	0.432	0.357	0.410
		6	0.331	0.375	0.212	0.432	0.407	0.374	0.347	0.438	0.370	0.451	0.398	0.417
		7	0.336	0.389	0.217	0.439	0.417	0.408	0.350	0.450	0.348	0.442	0.396	0.434
		8	0.319	0.382	0.215	0.453	0.387	0.389	0.347	0.475	0.379	0.439	0.402	0.407
		9	0.342	0.392	0.216	0.431	0.400	0.399	0.341	0.447	0.390	0.444	0.397	0.425
		10	0.324	0.391	0.215	0.445	0.423	0.409	0.378	0.460	0.360	0.438	0.371	0.416
3	F1_N3	1	0.326	0.405	0.227	0.456	0.418	0.393	0.372	0.476	0.433	0.439	0.399	0.440
		2	0.359	0.404	0.218	0.445	0.432	0.416	0.369	0.454	0.387	0.464	0.390	0.433
		3	0.330	0.391	0.218	0.434	0.402	0.390	0.370	0.462	0.362	0.439	0.360	0.436
		4	0.351	0.404	0.221	0.455	0.401	0.411	0.367	0.479	0.402	0.437	0.374	0.424
		5	0.315	0.391	0.210	0.438	0.429	0.395	0.362	0.459	0.384	0.440	0.369	0.406
		6	0.322	0.371	0.206	0.426	0.386	0.381	0.350	0.441	0.432	0.438	0.397	0.416
		7	0.319	0.385	0.216	0.441	0.407	0.408	0.360	0.448	0.374	0.443	0.362	0.430
		8	0.318	0.383	0.221	0.452	0.388	0.389	0.347	0.477	0.385	0.442	0.402	0.423
		9	0.315	0.398	0.208	0.429	0.379	0.406	0.349	0.449	0.401	0.442	0.362	0.430
		10	0.329	0.391	0.210	0.452	0.416	0.407	0.357	0.477	0.366	0.424	0.361	0.418
4	F1_N4	1	0.328	0.331	0.235	0.411	0.410	0.336	0.349	0.459	0.405	0.399	0.411	0.403
		2	0.341	0.320	0.220	0.412	0.422	0.331	0.371	0.430	0.387	0.390	0.385	0.384
		3	0.338	0.317	0.232	0.391	0.421	0.340	0.367	0.438	0.378	0.398	0.366	0.391
		4	0.331	0.331	0.211	0.410	0.401	0.346	0.367	0.445	0.406	0.380	0.355	0.413
		5	0.312	0.322	0.217	0.412	0.432	0.341	0.390	0.455	0.392	0.364	0.390	0.392
		6	0.322	0.319	0.210	0.406	0.392	0.336	0.345	0.414	0.408	0.394	0.394	0.415
		7	0.315	0.313	0.211	0.399	0.422	0.334	0.385	0.428	0.356	0.365	0.370	0.417
		8	0.311	0.316	0.221	0.433	0.381	0.333	0.358	0.455	0.401	0.377	0.388	0.400
		9	0.338	0.311	0.221	0.393	0.396	0.322	0.353	0.424	0.394	0.379	0.368	0.391
		10	0.326	0.315	0.212	0.403	0.407	0.329	0.337	0.437	0.377	0.366	0.360	0.422
5	F1_N5	1	0.316	0.412	0.239	0.357	0.401	0.405	0.364	0.449	0.467	0.420	0.410	0.426
		2	0.345	0.397	0.222	0.359	0.406	0.396	0.366	0.418	0.406	0.465	0.354	0.378
		3	0.371	0.397	0.223	0.352	0.423	0.386	0.348	0.409	0.399	0.460	0.368	0.397
		4	0.355	0.403	0.208	0.349	0.385	0.400	0.368	0.415	0.430	0.420	0.355	0.418
		5	0.313	0.405	0.211	0.345	0.366	0.404	0.370	0.407	0.403	0.415	0.408	0.376
		6	0.338	0.400	0.209	0.347	0.395	0.379	0.343	0.397	0.419	0.436	0.400	0.400
		7	0.338	0.389	0.223	0.345	0.389	0.403	0.374	0.399	0.406	0.417	0.385	0.395
		8	0.332	0.399	0.227	0.365	0.384	0.389	0.357	0.442	0.382	0.436	0.385	0.404
		9	0.310	0.405	0.211	0.351	0.394	0.399	0.365	0.396	0.411	0.431	0.389	0.387
		10	0.315	0.402	0.214	0.351	0.425	0.390	0.362	0.414	0.386	0.421	0.354	0.402
6	F2_N1	1	0.264	0.435	0.176	0.282	0.276	0.465	0.253	0.319	0.333	0.495	0.293	0.308
		2	0.257	0.442	0.160	0.272	0.305	0.456	0.254	0.317	0.296	0.480	0.262	0.304
		3	0.231	0.447	0.169	0.300	0.276	0.462	0.271	0.375	0.322	0.499	0.245	0.330

Rank Loss			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
		4	0.263	0.441	0.166	0.291	0.289	0.461	0.263	0.323	0.270	0.512	0.292	0.335
		5	0.269	0.442	0.169	0.297	0.294	0.463	0.250	0.378	0.275	0.487	0.267	0.342
		6	0.256	0.447	0.164	0.282	0.288	0.453	0.261	0.346	0.322	0.515	0.263	0.344
		7	0.253	0.439	0.173	0.279	0.263	0.466	0.235	0.343	0.281	0.497	0.257	0.330
		8	0.256	0.437	0.176	0.290	0.272	0.467	0.240	0.358	0.282	0.502	0.284	0.318
		9	0.226	0.441	0.167	0.284	0.247	0.465	0.212	0.360	0.292	0.516	0.251	0.319
		10	0.254	0.441	0.162	0.308	0.258	0.462	0.259	0.329	0.308	0.482	0.278	0.360
7	F2_N2	1	0.264	0.438	0.173	0.283	0.276	0.465	0.273	0.324	0.330	0.489	0.303	0.309
		2	0.256	0.442	0.169	0.272	0.305	0.456	0.273	0.311	0.282	0.483	0.246	0.313
		3	0.236	0.449	0.168	0.300	0.276	0.464	0.243	0.375	0.322	0.505	0.259	0.338
		4	0.267	0.439	0.159	0.292	0.289	0.460	0.245	0.323	0.268	0.512	0.278	0.341
		5	0.272	0.442	0.161	0.296	0.294	0.463	0.241	0.380	0.305	0.487	0.270	0.341
		6	0.257	0.447	0.160	0.283	0.290	0.453	0.248	0.349	0.319	0.515	0.277	0.339
		7	0.253	0.439	0.160	0.278	0.263	0.466	0.254	0.339	0.274	0.501	0.273	0.320
		8	0.250	0.449	0.179	0.290	0.272	0.473	0.241	0.358	0.279	0.500	0.290	0.327
		9	0.224	0.440	0.163	0.284	0.248	0.465	0.247	0.362	0.292	0.516	0.244	0.317
		10	0.254	0.441	0.167	0.306	0.257	0.462	0.268	0.329	0.307	0.482	0.280	0.367
8	F2_N3	1	0.264	0.435	0.175	0.288	0.276	0.465	0.257	0.326	0.330	0.489	0.302	0.309
		2	0.256	0.442	0.163	0.274	0.305	0.456	0.251	0.312	0.282	0.480	0.279	0.299
		3	0.236	0.449	0.169	0.296	0.276	0.462	0.243	0.377	0.325	0.505	0.268	0.331
		4	0.267	0.439	0.165	0.292	0.289	0.459	0.258	0.322	0.268	0.512	0.270	0.330
		5	0.272	0.442	0.163	0.296	0.295	0.463	0.248	0.382	0.308	0.493	0.265	0.347
		6	0.257	0.447	0.158	0.284	0.290	0.453	0.251	0.350	0.319	0.515	0.264	0.345
		7	0.253	0.439	0.164	0.278	0.263	0.466	0.234	0.339	0.274	0.497	0.268	0.308
		8	0.250	0.444	0.177	0.290	0.272	0.467	0.252	0.358	0.280	0.510	0.284	0.312
		9	0.224	0.440	0.162	0.286	0.247	0.465	0.226	0.368	0.292	0.509	0.265	0.312
		10	0.254	0.441	0.168	0.309	0.258	0.462	0.265	0.333	0.307	0.495	0.286	0.336
9	F2_N4	1	0.264	0.314	0.175	0.213	0.276	0.345	0.275	0.297	0.330	0.352	0.284	0.258
		2	0.256	0.306	0.162	0.209	0.305	0.334	0.257	0.291	0.282	0.315	0.297	0.254
		3	0.236	0.314	0.165	0.232	0.276	0.330	0.258	0.316	0.325	0.332	0.263	0.273
		4	0.267	0.312	0.166	0.207	0.289	0.337	0.251	0.308	0.268	0.350	0.288	0.263
		5	0.272	0.341	0.177	0.225	0.295	0.352	0.235	0.343	0.308	0.341	0.277	0.263
		6	0.257	0.307	0.163	0.203	0.290	0.324	0.262	0.316	0.319	0.338	0.267	0.259
		7	0.251	0.282	0.167	0.209	0.263	0.322	0.234	0.304	0.274	0.323	0.261	0.248
		8	0.250	0.307	0.176	0.222	0.272	0.339	0.253	0.317	0.280	0.357	0.293	0.278
		9	0.224	0.318	0.163	0.226	0.247	0.342	0.238	0.305	0.292	0.344	0.272	0.259
		10	0.254	0.307	0.169	0.222	0.258	0.342	0.245	0.322	0.307	0.344	0.290	0.272
10	F2_N5	1	0.263	0.433	0.172	0.231	0.273	0.458	0.268	0.312	0.335	0.498	0.295	0.256
		2	0.257	0.429	0.169	0.235	0.314	0.455	0.260	0.307	0.296	0.498	0.261	0.242
		3	0.231	0.428	0.168	0.236	0.276	0.447	0.256	0.324	0.324	0.516	0.264	0.262
		4	0.263	0.422	0.163	0.236	0.286	0.452	0.246	0.340	0.278	0.513	0.251	0.276
		5	0.269	0.432	0.168	0.240	0.295	0.453	0.238	0.345	0.276	0.483	0.281	0.285
		6	0.255	0.436	0.164	0.233	0.288	0.450	0.253	0.316	0.319	0.508	0.267	0.279
		7	0.252	0.427	0.164	0.223	0.263	0.447	0.251	0.302	0.283	0.508	0.253	0.254
		8	0.262	0.436	0.179	0.230	0.274	0.459	0.239	0.329	0.282	0.503	0.290	0.302
		9	0.223	0.431	0.157	0.249	0.247	0.455	0.239	0.322	0.297	0.528	0.255	0.282
		10	0.254	0.433	0.167	0.245	0.256	0.457	0.236	0.316	0.311	0.505	0.289	0.296

Lampiran 7. Hasil *Average Precision* pada setiap dataset

Avg Precision			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
1	F1_N1	1	0.597	0.470	0.715	0.501	0.501	0.548	0.527	0.388	0.499	0.555	0.541	0.452
		2	0.622	0.481	0.715	0.527	0.513	0.535	0.553	0.395	0.500	0.539	0.564	0.461
		3	0.580	0.501	0.702	0.533	0.473	0.556	0.543	0.407	0.501	0.558	0.566	0.464
		4	0.620	0.480	0.713	0.507	0.526	0.543	0.531	0.378	0.527	0.559	0.544	0.452
		5	0.672	0.499	0.736	0.529	0.550	0.553	0.549	0.398	0.516	0.553	0.577	0.468
		6	0.595	0.505	0.727	0.528	0.454	0.564	0.567	0.414	0.491	0.557	0.556	0.465
		7	0.623	0.498	0.704	0.525	0.523	0.543	0.539	0.407	0.516	0.551	0.571	0.449
		8	0.633	0.505	0.713	0.518	0.512	0.561	0.545	0.406	0.478	0.556	0.553	0.473
		9	0.639	0.488	0.739	0.525	0.490	0.553	0.564	0.405	0.501	0.548	0.565	0.471
		10	0.604	0.494	0.722	0.515	0.494	0.542	0.550	0.398	0.488	0.558	0.578	0.471
2	F1_N2	1	0.612	0.473	0.692	0.509	0.501	0.551	0.523	0.390	0.474	0.556	0.545	0.472
		2	0.615	0.471	0.711	0.528	0.497	0.537	0.535	0.399	0.547	0.539	0.563	0.468
		3	0.603	0.494	0.708	0.538	0.499	0.550	0.540	0.412	0.535	0.562	0.574	0.467
		4	0.608	0.481	0.726	0.505	0.498	0.540	0.571	0.378	0.483	0.556	0.559	0.472
		5	0.628	0.498	0.732	0.533	0.500	0.553	0.563	0.401	0.519	0.558	0.576	0.479
		6	0.620	0.503	0.734	0.531	0.504	0.567	0.562	0.431	0.533	0.546	0.553	0.466
		7	0.611	0.495	0.705	0.529	0.501	0.543	0.554	0.406	0.553	0.549	0.552	0.452
		8	0.647	0.502	0.732	0.526	0.524	0.563	0.561	0.408	0.530	0.552	0.549	0.483
		9	0.596	0.493	0.722	0.529	0.531	0.554	0.567	0.411	0.532	0.549	0.548	0.463
		10	0.604	0.494	0.720	0.517	0.492	0.543	0.537	0.400	0.526	0.555	0.566	0.471
3	F1_N3	1	0.604	0.476	0.700	0.506	0.502	0.557	0.542	0.391	0.469	0.550	0.551	0.458
		2	0.601	0.479	0.712	0.527	0.487	0.530	0.544	0.400	0.521	0.536	0.547	0.464
		3	0.610	0.498	0.719	0.533	0.497	0.558	0.529	0.408	0.523	0.558	0.567	0.456
		4	0.606	0.484	0.708	0.502	0.508	0.540	0.550	0.379	0.489	0.555	0.573	0.470
		5	0.630	0.498	0.729	0.528	0.490	0.553	0.559	0.402	0.527	0.552	0.565	0.480
		6	0.623	0.502	0.733	0.532	0.529	0.562	0.555	0.427	0.481	0.556	0.544	0.469
		7	0.617	0.500	0.708	0.527	0.491	0.543	0.546	0.410	0.527	0.552	0.580	0.461
		8	0.636	0.503	0.717	0.520	0.526	0.563	0.571	0.403	0.509	0.554	0.540	0.468
		9	0.617	0.488	0.724	0.523	0.525	0.548	0.562	0.402	0.515	0.546	0.571	0.468
		10	0.599	0.494	0.713	0.509	0.488	0.545	0.548	0.395	0.513	0.564	0.572	0.469
4	F1_N4	1	0.613	0.557	0.696	0.550	0.511	0.595	0.559	0.430	0.498	0.557	0.525	0.480
		2	0.616	0.565	0.711	0.560	0.494	0.589	0.545	0.428	0.533	0.559	0.551	0.498
		3	0.604	0.578	0.697	0.582	0.484	0.586	0.547	0.447	0.523	0.562	0.577	0.506
		4	0.615	0.559	0.725	0.553	0.511	0.589	0.549	0.433	0.488	0.568	0.578	0.486
		5	0.618	0.563	0.713	0.560	0.480	0.594	0.538	0.440	0.518	0.588	0.551	0.493
		6	0.617	0.575	0.722	0.566	0.533	0.589	0.563	0.451	0.506	0.557	0.536	0.486
		7	0.620	0.577	0.704	0.561	0.487	0.594	0.535	0.432	0.549	0.583	0.555	0.471
		8	0.647	0.564	0.723	0.545	0.520	0.601	0.567	0.436	0.513	0.574	0.550	0.486
		9	0.596	0.574	0.711	0.572	0.520	0.603	0.551	0.447	0.539	0.581	0.565	0.497
		10	0.597	0.569	0.721	0.563	0.503	0.603	0.567	0.441	0.516	0.590	0.562	0.476
5	F1_N5	1	0.617	0.477	0.697	0.591	0.517	0.549	0.552	0.443	0.441	0.560	0.531	0.481
		2	0.623	0.492	0.704	0.593	0.494	0.554	0.549	0.457	0.486	0.530	0.571	0.507
		3	0.582	0.487	0.721	0.602	0.474	0.565	0.568	0.472	0.511	0.539	0.571	0.514
		4	0.614	0.481	0.727	0.592	0.508	0.554	0.548	0.465	0.487	0.572	0.580	0.486
		5	0.634	0.482	0.717	0.604	0.538	0.545	0.543	0.466	0.497	0.570	0.545	0.513
		6	0.619	0.490	0.729	0.602	0.501	0.565	0.562	0.462	0.477	0.548	0.535	0.501
		7	0.604	0.497	0.705	0.601	0.516	0.549	0.533	0.482	0.494	0.569	0.560	0.504
		8	0.619	0.488	0.710	0.581	0.524	0.566	0.555	0.451	0.476	0.558	0.550	0.497
		9	0.630	0.472	0.726	0.606	0.524	0.549	0.548	0.483	0.516	0.557	0.559	0.512
		10	0.619	0.477	0.714	0.605	0.491	0.561	0.547	0.469	0.506	0.570	0.578	0.496
6	F2_N1	1	0.697	0.492	0.775	0.661	0.616	0.493	0.656	0.491	0.542	0.521	0.612	0.548
		2	0.716	0.483	0.791	0.654	0.595	0.501	0.653	0.496	0.575	0.531	0.646	0.550
		3	0.738	0.476	0.784	0.633	0.617	0.496	0.647	0.467	0.564	0.520	0.661	0.527

Avg Precision			BR				CC				LC			
No.	Dataset	RS	J48	SMO	RF	NB	J48	SMO	RF	NB	J48	SMO	RF	NB
		4	0.704	0.491	0.785	0.636	0.612	0.496	0.634	0.492	0.597	0.507	0.612	0.523
		5	0.713	0.487	0.790	0.635	0.609	0.496	0.645	0.463	0.605	0.532	0.636	0.505
		6	0.711	0.481	0.783	0.651	0.606	0.504	0.642	0.498	0.559	0.504	0.647	0.518
		7	0.696	0.492	0.785	0.655	0.626	0.487	0.657	0.493	0.610	0.523	0.642	0.512
		8	0.700	0.491	0.776	0.643	0.619	0.489	0.653	0.477	0.584	0.509	0.607	0.526
		9	0.740	0.484	0.796	0.654	0.660	0.493	0.683	0.469	0.568	0.508	0.646	0.524
		10	0.723	0.487	0.790	0.629	0.633	0.495	0.647	0.488	0.584	0.533	0.633	0.507
7	F2_N2	1	0.697	0.489	0.778	0.656	0.616	0.493	0.627	0.490	0.550	0.527	0.599	0.547
		2	0.717	0.483	0.786	0.654	0.595	0.502	0.636	0.501	0.592	0.530	0.642	0.546
		3	0.735	0.477	0.786	0.633	0.617	0.494	0.660	0.467	0.564	0.514	0.637	0.515
		4	0.698	0.493	0.780	0.635	0.612	0.496	0.654	0.491	0.607	0.507	0.628	0.524
		5	0.710	0.487	0.786	0.637	0.610	0.496	0.664	0.464	0.580	0.532	0.641	0.505
		6	0.711	0.481	0.782	0.650	0.603	0.505	0.653	0.497	0.563	0.504	0.629	0.523
		7	0.696	0.492	0.782	0.654	0.626	0.487	0.647	0.493	0.617	0.520	0.632	0.521
		8	0.702	0.478	0.776	0.643	0.619	0.483	0.653	0.477	0.583	0.513	0.605	0.517
		9	0.743	0.486	0.792	0.652	0.657	0.494	0.665	0.467	0.568	0.508	0.654	0.526
		10	0.723	0.487	0.791	0.634	0.633	0.495	0.630	0.485	0.586	0.533	0.633	0.501
8	F2_N3	1	0.697	0.492	0.772	0.649	0.616	0.493	0.647	0.496	0.550	0.527	0.609	0.550
		2	0.717	0.483	0.789	0.654	0.595	0.502	0.644	0.498	0.592	0.531	0.630	0.561
		3	0.735	0.474	0.782	0.636	0.617	0.496	0.657	0.467	0.560	0.514	0.637	0.517
		4	0.698	0.492	0.782	0.632	0.612	0.498	0.636	0.496	0.607	0.507	0.625	0.531
		5	0.710	0.487	0.795	0.635	0.609	0.496	0.653	0.462	0.577	0.526	0.633	0.501
		6	0.711	0.481	0.788	0.646	0.603	0.504	0.650	0.497	0.563	0.504	0.643	0.521
		7	0.696	0.492	0.793	0.655	0.625	0.487	0.661	0.498	0.617	0.523	0.640	0.527
		8	0.702	0.483	0.775	0.643	0.619	0.489	0.645	0.476	0.582	0.506	0.603	0.527
		9	0.743	0.486	0.802	0.651	0.660	0.492	0.677	0.466	0.568	0.510	0.647	0.537
		10	0.723	0.487	0.795	0.631	0.633	0.495	0.645	0.493	0.586	0.525	0.629	0.527
9	F2_N4	1	0.697	0.598	0.769	0.742	0.616	0.599	0.634	0.580	0.550	0.608	0.611	0.589
		2	0.717	0.598	0.792	0.739	0.595	0.598	0.646	0.581	0.592	0.638	0.611	0.608
		3	0.735	0.588	0.789	0.723	0.617	0.606	0.646	0.573	0.560	0.621	0.634	0.574
		4	0.698	0.597	0.780	0.744	0.612	0.599	0.647	0.570	0.607	0.614	0.623	0.595
		5	0.710	0.572	0.769	0.729	0.609	0.592	0.668	0.550	0.577	0.619	0.641	0.583
		6	0.711	0.598	0.787	0.748	0.603	0.615	0.644	0.578	0.563	0.623	0.648	0.595
		7	0.699	0.620	0.789	0.750	0.625	0.616	0.662	0.589	0.617	0.632	0.641	0.592
		8	0.702	0.599	0.779	0.733	0.619	0.602	0.644	0.574	0.582	0.601	0.596	0.566
		9	0.743	0.590	0.791	0.722	0.660	0.594	0.661	0.573	0.568	0.614	0.642	0.586
		10	0.723	0.600	0.789	0.731	0.633	0.590	0.654	0.565	0.586	0.606	0.621	0.589
10	F2_N5	1	0.697	0.493	0.776	0.710	0.618	0.495	0.634	0.537	0.541	0.518	0.618	0.592
		2	0.716	0.496	0.785	0.699	0.587	0.499	0.634	0.534	0.574	0.520	0.642	0.602
		3	0.738	0.497	0.783	0.698	0.617	0.506	0.650	0.538	0.561	0.506	0.641	0.581
		4	0.704	0.502	0.780	0.704	0.614	0.502	0.650	0.526	0.590	0.510	0.650	0.574
		5	0.713	0.495	0.786	0.698	0.609	0.505	0.659	0.522	0.607	0.535	0.633	0.564
		6	0.712	0.494	0.783	0.705	0.606	0.507	0.650	0.538	0.562	0.511	0.644	0.568
		7	0.696	0.496	0.788	0.724	0.626	0.503	0.646	0.546	0.609	0.514	0.646	0.582
		8	0.698	0.489	0.781	0.713	0.619	0.492	0.653	0.527	0.579	0.508	0.612	0.548
		9	0.744	0.493	0.798	0.689	0.660	0.502	0.653	0.526	0.567	0.494	0.649	0.561
		10	0.723	0.494	0.787	0.691	0.634	0.494	0.671	0.529	0.583	0.515	0.628	0.568



## BIOGRAFI PENULIS



**Allantutra Guslaw**, Magister di Bidang Keahlian Telematika Konsentrasi Pengelola TIK Pemerintahan (PeTIK), Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya angkatan 2016. Merupakan salah satu penerima Beasiswa Kerjasama Kementerian Komunikasi dan Informasi Republik Indonesia dengan Badan Pemeriksa Keuangan Republik Indonesia. Lahir di Padang pada tanggal 09 Agustus 1982. Anak pertama dari 4 bersaudara dari pasangan (alm) Buyung Adik dan Murni A., suami dari Gusti Limbak Cahayo dan ayah dari Amirah Ghassani. Bertugas pada Badan Pemeriksa Keuangan Republik Indonesia.

Alamat Email: [allan.guslaw@gmail.com](mailto:allan.guslaw@gmail.com)

### **Riwayat Pendidikan:**

1988 - 1994 : SD Yos Sudarso Padang, Sumatera Barat

1994 - 1997 : SMP Negeri 12 Padang, Sumatera Barat

1997 - 2000 : SMU Negeri 3 Padang, Sumatera Barat

2000 – 2005 : Teknik Informatika, Universitas Gunadarma, Depok

### **Riwayat Pekerjaan:**

2007 - 2008 : PT. Adira Dinamika Multi Finance Cabang Padang

2008 – sekarang : Badan Pemeriksa Keuangan Republik Indonesia